2005

Identification of 'structural alerts' and associated mechanisms of action of mammary gland carcinogens in female rodents

Shanna Tommika Moss
Louisiana State University and Agricultural and Mechanical College

IDENTIFICATION OF 'STRUCTURAL ALERTS' AND ASSOCIATED MECHANISMS OF ACTION OF MAMMARY GLAND CARCINOGENS IN FEMALE RODENTS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science

in

The Department of Environmental Studies

by
Shanna T. Moss
B.S., University of Nebraska, 2002
May 2005

ACKNOWLEDGEMENTS

I gratefully acknowledge support for this work from the Congressionally Directed Medical Research Program for Breast Cancer Idea Award DAMD17-01-0376. Moreover, several persons provided guidance and support for the successful completion of this project from the very beginning. My most sincere gratitude goes out to my very considerate research advisor, Dr. Albert R. Cunningham, whose impressive research work sparked my interest in computational toxicology. I am very grateful for the time and patience he allotted to help get this massive, and I think we both can agree on, sometimes tedious research completed in a timely fashion. Thanks for the well-equipped lab Dr. Cunningham.

Furthermore, I would like to thank Suzanne L. Cunningham for her valuable technical contribution to this project. I want to thank my committee members Dr. Vincent L. Wilson and Dr. Ralph J. Portier for serving on my committee and also for their scientific input and willingness to assist whenever they can. I am also grateful for their editorial assistance along the way. I am cordially thankful to my colleague Daniel M. Consoer for allowing me to use his ESCREEN models for statistical evaluation. I truly enjoyed our conversations at the lab.

I would like to thank my parents, for their continuous moral and financial support throughout the years. Special thanks to my brother and sister, Duran and Kay, for their encouragement, and other supportive family members and friends I failed to mention. I want to thank my very caring boyfriend, B.J. Alexander, for his constant support and love through all of this. Thanks for keeping me up with the coffee B.J.

I would also like to extend my most heartfelt thanks and appreciation to Ms. Charlotte St. Romain who has been there for me since day one. Thanks for keeping me on track and for your guidance Charlotte. Finally, I want to thank all the faculty and friends I have

encountered in the Department of Environmental Studies who have enlightened me with their knowledge and long-lasting friendship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CASE	Computer Automated Structure Evaluation
Cat-SAR	Categorical Structure-Activity Relationship
CDA	Chemical Diversity Approach
CDMRP	Congressionally Directed Medical Research Program
CPDB	Carcinogenic Potency Database
DEREK	Deductive Estimation of Risk from Existing Knowledge
DES	Diethylstilbestrol
DNA	Deoxyribonucleic Acid
EDC	Endocrine-Disrupting Chemicals
EPA	Environmental Protection Agency
ER	Estrogen Receptor
ERK	Extracellular Signal-Regulated Kinase
HPVC	High-Production Volume Chemical
HQSAR	Holographic Quantitative Structure-Activity Relationship
HRT	Hormone Replacement Therapy
LOO-CV	Leave-One-Out Cross-validation
MC	Mammary Carcinogen
MTD	Maximum Tolerated Dose
NC	Noncarcinogen
NCI	National Cancer Institute
NMC	Non-Mammary Carcinogen
NTP	National Toxicology Program

OCP	Observed Correct Prediction
PAH	Polynuclear Aromatic Hydrocarbon
PCB	Polychlorinated Biphenyl
PMN	Premanufacturing Notice
RNC	Rodent Noncarcinogen
RPE	Relative Proliferative Effect
RPP	Relative Proliferative Potency
SAR	Structure-Activity Relationship
TD_{50}	Toxic Dose (at which 50% of test animals die)
TOPKAT	Toxicity Prediction by Computer Assisted Technology
QSAR	Quantitative Structure-Activity Relationship

ABSTRACT

A new structure-activity relationship (SAR) approach to modeling was utilized to study mammary gland carcinogens. A set of chemicals tested for mammary tumorigenesis that have been analyzed in the Carcinogenic Potency Database (CPDB) were subjected to several computational analyses in an attempt to predict each chemical's carcinogenic potential. A total of six learning sets (rat and mouse mammary gland carcinogen, CPDB rat and mouse, and female-specific rodent models) were developed and validated using a SAR modeling algorithm called categorical-SAR (cat-SAR). The predictive cat-SAR program evaluates active and inactive compounds of known biological activity and predicts their biological activity from this categorical data. Overall, this study demonstrates the usefulness of cat-SAR and its successful application in developing 'structural alerts' to breast carcinogenicity. The resulting rat and mouse mammary carcinogen models achieved an 82.0% (sensitivity 76.7%; specificity 87.5%) and 80.6% (sensitivity 80%; specificity 81.8%) concordance between experimental and predicted results, respectively. Likewise, the general CPDB mouse and rat models were both 70% predictive. Corresponding sensitivity and specificity values were 74.2 and 66.7% and 70.4 and 68.5%, respectively. The analyses indicate the capability of cat-SAR in identifying molecular fragments that potentially interact with cellular components present only in the targeted cell type (e.g., breast tissue cells). Moreover, this method is expected to help pre-determine structural alerts to carcinogen-induced mammary cancer. Identification of these 'structural alerts' can assist in understanding mechanisms involved in making a normal breast cell cancerous. Using the results of these analyses, it is possible to classify and rank structurally diverse chemicals as to their potential to induce mammary gland cancer.

CHAPTER 1. INTRODUCTION

According to the National Cancer Institute (NCI), about one in seven women in America will be diagnosed with some form of breast cancer during her lifetime, a rate that has increased over the last six decades from 1 in 22 (Ries et al.; NCI 2004). Breast cancer when compared to other types of cancer is the most diagnosed and prevalent cancer in women worldwide. Unfortunately, since the rate is increasing it has been estimated that breast cancer will account for 28% of all female cancers by 2012 (Hodgen et al 2002). The breast cancer genes, BRCA1 and BRCA2, account for only five percent of all breast cancer cases (Davis et al 1995). Furthermore, female BRCA1 or BRCA2 mutation carriers have a lifetime risk of breast cancer of between 50% and 80% (Martin et al 2000). While ionizing radiation, diet, nulliparity, smoking, late childbirth, and genetic factors account for at best 47% of cases (Madigan et al 1995), there is reason to speculate that exposure to industrial (e.g., organochlorines and pesticides), pharmaceutical, and some plant products (e.g., phytoestrogens) may play a role in the rising number of breast cancer cases (Davis et al 1993). An upper estimated limit of 80% of all cancer cases has been pinpointed to the environment (Tomatis et al 2001). However, it is important to note that finding plausible direct links between any type of cancer and the environment is complex.

Additionally, the Congressionally Directed Medical Research Program (CDMRP) statistics reports that there are 192,000 established new cases of breast cancer each year, resulting in 40,200 deaths (CDMRP 2004). The current cycle of funding includes more than $150 million for breast cancer research. Also according to the CDMRP, the American Cancer Society figures for 2004 show that approximately 216,000 women in the United States are projected to receive a diagnosis of invasive breast cancer. Thus, the assessment of breast

cancer is a major public health concern that must be given immediate attention. Hence, this study is being conducted for several important reasons.

First of all, current methods of high-dose animal cancer bioassays do not provide enough information to assess human cancer risks at the usual levels of exposure (Gold et al 2002). Humans are exposed to a myriad of potentially harmful environmental agents at relatively low doses. Hence, a cost-effective and rapid tool may be useful in the screening and prioritization of environmental chemicals (e.g., commercial products such as pesticides, cosmetic ingredients, nutritional supplements, and food additives). These include high-production volume chemicals (HPVCs) that are produced or imported in quantities exceeding 1 million pounds per year. The United States Environmental Protection Agency (U.S. EPA) has noted 87,000 chemicals to be screened for endocrine disruption and other toxicological phenomena (EPA 2002). Furthermore, a typical drug takes ten to twelve years and costs up to $500 million to reach the market, it is clearly important to discover potential toxicity *in silico* as soon as possible (Dearden 2003). By establishing an estimated risk posed by groups of environmental chemicals it may be possible to control their exposure or to facilitate the synthesis of safer consumer products.

Secondly, the study presented herein will attempt to predict the fate and effects of such chemicals on human and animal health via a new computational method called categorical-SAR (cat-SAR). In addition to developing highly accurate models, the cat-SAR program presents a better explanatory power in comparison to other SAR approaches. The final suitably validated SAR models of this study may have significant utility as a two-dimensional screening tool, to the extent that large datasets of untested candidate compounds

can now be rapidly assessed and prioritized for subsequent toxicological evaluation as mammary carcinogens.

Thirdly, it has been hypothesized that endocrine-disrupting chemicals (EDCs), particularly, synthetic estrogenic environmental contaminants (e.g., xenoestrogens) are linked to an increased incidence of breast cancer in women. EDCs (e.g., bisphenol A and 4-nonylphenol) mimic the action of the female steroid hormone 17-β-estradiol, which is primarily responsible for the development of the female reproductive system (Crisp et al 1998). However, there have been conflicting epidemiological data regarding the role of xenoestrogens in breast cancer development (Safe 1995; Falck 1992; Ashby 1997). Most research studies do not support the hypothesis that there is a link between exposure to environmental chemicals, particularly endocrine-disrupting chemicals (EDCs), and the increasing incidence of breast cancer (Safe 2004). Hence, the underlying relationship between carcinogens and estrogens was investigated.

Governmental and public concern over the high prevalence of structurally diverse chemicals some of which are toxic to humans, especially chemicals present in consumer and industrial products, medicines, foods, the workplace, and the environment has now surfaced as one of the most pressing issues of interest. The National Toxicology Program (NTP) was established in 1978 to coordinate research and testing of potential human carcinogens and to publish the Report on Carcinogens, which lists human carcinogens (Bennett and Davis, 2002). SAR models have been accepted by regulatory agencies worldwide, including the U.S. EPA, for risk assessment. Predictive and mechanistic SAR models have shown to be highly reliable in terms of assessing untested chemicals for toxic potential and understanding how they may induce toxicity.

For this study, an attempt was made to yield the best concordance between predictions and experimental results. It is proposed that the approach described herein will curtail the stupendous amount of time, cost, and use of animals involved in bioassays to assess these compounds on a regulatory basis. The SAR analysis should thereby provide further insight into the extent (if at all) xenoestrogens potentially play in the etiology of breast cancer. The hypothesis for this study is that specific attributes of certain chemicals are related to their ability to induce breast cancer. In summary, the goals and objectives of this study are as follow:

(i) Evaluate a new SAR technique called cat-SAR in identifying structural alerts of carcinogenesis.

(ii) Identify structural alerts for rodent and mammary-specific carcinogenesis.

(iii) Investigate the mechanistic underpinnings of mutagenesis and estrogenicity and their potential involvement in breast cancer development.

The stated goals were achieved via use of the cat-SAR program in conjunction with Tripos Sybyl Holographic Quantitative Structure-Activity Relationship (HQSAR) modeling software. Of importance, the approach described herein helps determine relatedness (i.e., structural overlap) among different toxicological endpoints such as rodent carcinogenesis, female and breast-specific carcinogenesis, estrogenicity, and mutagenesis by examining a random sample of 10,000 chemicals of unknown biological activity. This group of compounds derived from chemical structure libraries and from a random sample of chemical structures from the NCI is representative of all chemicals in the environment. The prevalence of chemicals predicted to possess the ability to induce multiple biological effects

simultaneously should provide a measure of the mechanistic relatedness of these toxicological phenomena.

CHAPTER 2. LITERATURE REVIEW

2.1 Breast Cancer Development

Breast cancer is a broad term used to describe various types of cancer that occur in the breast. Like most other common and life-threatening epithelial malignancies, human breast cancer presents itself clinically after a prolonged multi-stage process of carcinogenesis. Breast tumors are most likely to arise in the epithelial cells that line the mammary glands and ducts and are associated with several carcinogen exposures including radiation, diethylstilbestrol (DES), and estrogens (Brown et al 1995). Mammary gland growth and development are mediated through the complex interactions of steroid hormones, polypeptide hormones, growth stimulatory factors, and growth inhibitory factors (Haslam et al 2003).

Breast cancer is very uncommon in women under the age of thirty-five, but it is prevalent in women over the age of fifty, and the risk is especially high for women over age sixty (Ganz 2001). Ettinger et al have proposed that women (usually women over age 35) using estrogen replacement therapy have a higher risk of breast cancer, although the levels do not reach statistical significance (Ettinger et al 1996). Moreover, it is evident that long-term users of hormone replacement therapy (HRT) and those exposed to high doses of estrogen may slightly increase the risk of breast cancer ((Brinton et al 1993). In support of this, a recent study presented findings that show a hyperproliferative effect of HRT on mammary epithelial cells (Le Marchand 2004). Some researchers have reported a greater risk of breast cancer in postmenopausal women receiving combined estrogen plus progestin HRT than in those receiving estrogen alone that may indicate a significant role for progesterone in

mammary cancer (Stadel 2002). However, little is known about risk from the frequently prescribed estrogen/progestin combination.

Despite ongoing human epidemiological studies, the exact causes of breast cancer remain unknown. This is due to the complexity of carcinogenesis and the endocrine system. The endocrine system is composed of a number of elements (e.g., receptors, chemical messengers, and the synthetic apparatus) and actions (e.g., transcriptional (i.e., genomic) and non-transcriptional (i.e., non-genomic) mechanisms of signal transduction through steroid hormone receptors, homeostatic mechanisms, etc.)), all of which may play a role in the development of breast cancer (Simoncini et al 2003). Additionally, it has been presumed that certain structural attributes of estrogens are associated with metabolic transformation to mutagens that are highly engaged in mitogenesis (Davis et al 1993). Mitogenesis (i.e., cell proliferation) is one of the many mechanisms of cancer-causation. Estrogens have the natural ability to cause mitogenesis which makes it favorable to try to relate xenoestrogen exposure to breast cancer (Davis et al1993). However, xenoestrogens also have the potential to induce mammary cancer by exerting genotoxic or epigenetic effects on DNA via their possible biotransformation to free radicals (Liehr et al 1990). Roy and others have shown an elevated level of 8-hydroxyguanine through free radical generation of the prototype carcinogenic estrogen DES (Roy et al 1991).

Primary carcinogenesis may also occur with other xenobiotics because human breast epithelial and fibroblastic cells metabolize them to carcinogenic electrophiles (Davis et al 1993). For example, estradiol metabolism proceeds primarily through two mutually exclusive pathways, each of which is affected by xenobiotic exposures: Pathway I to 2-hydroxyestrone (2-OHE1), which has minimal estrogenic activity and is non-genotoxic, or Pathway II to 1-

alpha-hydroxyestrone (1-α-OHE1), a fully potent estrogen which is also genotoxic (Davis et al 1993). Breast cancer risk appears to be linked with these two pathways. This strengthens the reasoning behind attempting to link estrogenicity to breast carcinogenicity. Alterations of mammary gland development and to cellular proliferation from chemical exposure could alter cancer susceptibility (Brown et al 1995). To understand cancer it is important to know what happens when and how normal cells become cancerous, especially since it has already been noted that a dividing cell is much more at risk for mutation than a quiescent cell (Huff et al 1991).

2.2 Steroidal, Non-steroidal, and Synthetic Estrogens

It is noteworthy to introduce and review the different types of estrogens that humans are faced with as a large number of chemicals possess estrogenic activity and therefore, are possibly involved in breast cancer development. Although synthetic estrogens may be linked to breast cancer in women, there is a need for natural steroidal estrogen in female sexual maturation and growth (e.g., alter the distribution of body fat to produce body contour and other secondary characteristics). Additionally, high levels in some cells bring about pigmentation in the region of the nipples and areolae and in the genital region (Katzung 2004). Many steroidal estrogens are known as human carcinogens and are a class of hormonally active compounds derived from cholesterol with the primary purpose to control reproductive function and growth characteristics (IARC 1997). About 40% of all cancers in women are hormonally mediated (Henderson et al 1991). It has been proposed that these agents are associated with an increased breast cancer incidence in both women and men in the industrialized world (Katzung 2004).

In addition to the variety of steroidal estrogens derived from animal sources, numerous non-steroidal estrogens have been synthesized. Many phenols are estrogenic and estrogenic chemicals have been identified from diverse forms of life including those found in ocean sediments. Estrogen-mimetic compounds (bioflavonoids) are found in many plants as well, including saw palmetto and soybeans and other foods (Setchell 1985). The plant bioflavonoids include different structural classes of compounds, which contain a flavonoid backbone: flavones, flavanones, flavonols, isoflavones, and related condensation products (e.g., coumesterol) (Safe 1995).

An example of a non-steroidal estrogen, albeit pharmaceutical or antiestrogenic, is the most extensively used breast cancer drug, tamoxifen. Tamoxifen is a competitive partial agonist inhibitor of estradiol at the estrogen receptor used in the treatment of advanced breast cancer in postmenopausal women. Another synthetic and highly potent non-steroidal estrogen and known human carcinogen is DES, which was distributed to some 2 to 3 million women experiencing complications with their pregnancies between 1943 and 1971. As a result, hundreds of female offspring of women treated with DES have been diagnosed with clear cell adenocarcinoma (i.e., transplacental carcinogenesis) (Katzung 2004).

Some synthetic chemicals or intermediates that have been identified as estrogenic compounds include bisphenol A, a chemical used in the manufacture of polycarbonate-derived products, contaminants of phenol red solution, a pH indicator used in cell culture media, and alkyl phenols and their derivatives, which are extensively used for preparation of polyethoxylates in detergents (Safe 1995). Polychlorinated biphenyls (PCBs) and 1,1-dichloro-2,2-bis(*p*-chlorophenyl)ethylene (DDE) (a by-product of the pesticide DDT) are the two most researched organochlorine pollutants identified in all human tissues with high

frequencies. It should be noted that several hydroxylated PCBs bind to the estrogen receptor (ER), and it is possible that *para*-hydroxylated PCB metabolites may be the active estrogenic compounds associated with lower chlorinated PCBs. Some studies have reported that several additional organochlorine pesticides including endosulfan, toxaphene, and dieldrin exhibit estrogenlike activity and induce proliferation of estrogen responsive MCF-7 human breast cancer cells (Safe 1995).

2.3 Probable Modes of Action

There exist several 'structural alerts' or functional groups that may be associated with chemical mutagenesis and carcinogenesis. For carcinogenicity in humans, it is known that the vast majority of chemicals that cause cancer in humans are genotoxic, meaning that they attack DNA and pose as potent carcinogens. In contrast to human carcinogens, which are primarily organ-specific, most genotoxic rodent carcinogens are not species-specific or tissue-specific. It is noteworthy, however, that these chemicals represent the potential for either entering directly into electrophilic reaction with deoxyribonucleic acid (DNA) or being biotransformed by metabolism into an electrophilic functionality that can react with DNA (Ashby et al 1989). Increasing attention is now being given to the proven genotoxic and carcinogenic potential of endogenous estrogen metabolites especially 4-hydroxylation of estrogen to 4-hydroxy estrone (4-OH-E), via the cytochrome P450 enzyme, CYP1B1, which can form within and subsequently transform mammary epithelial cells independent of their estrogen receptor (ER) status (Butterworth et al 1992). Estrogens exert their physiological effects via the ER, which functions as a ligand-activated transcriptional regulator. ER is a member of a large family of nuclear receptor transcription factors with a characteristic modular structural organization with distinct domains associated with transactivation, DNA

binding and hormone binding. It is also an important pharmaceutical target for hormone replacement in menopausal women and for chemotherapeutic drugs against certain reproductive cancers (Pike et al 1999).

As described by Davis and others, "certain xenoestrogens, broadly defined in the literature as being basically hormone-mimicking compounds, may promote cancer by enhancing the production of "bad" estrogens. Other xenoestrogens may act by binding to the estrogen receptor and inducing it to issue unneeded proliferative signals." Chemicals with these properties may encourage the development of cancer in additional ways as well. For example, "there are indications that some xenoestrogens promote angiogenesis thus providing the blood vessels needed for tumor growth and spread (Davis et al 1993); others seem to damage DNA (i.e., mutagens)" (Davis et al 1995). Exposure at certain times may also heighten the carcinogenic effects of xenoestrogens (Davis and Bradlow 1993).

In addition to lifelong exposure to endogenous estrogen produced by the ovaries in pre-menopausal years and by aromatase conversion of androgens in post-menopausal years, susceptible breast epithelium may also be exposed to additive promoting doses of exogenous estrogen in the form of HRT, or to the as yet unproven *in vivo* promoting effects of environmental xenoestrogens (Davis et al 1993). Of importance, xenoestrogens are a relatively diverse group of compounds that do not lend themselves readily to typical SAR analysis.

2.4 Rodent Cancer Bioassay

Although, a large variety of compounds currently in general use are analogs to known mammary carcinogens for rodents, few have been tested for their carcinogenic potential (Coyle 2004). Approximately half of all chemicals tested, whether synthetic or natural, with

animal bioassays are carcinogenic to rats or mice at relatively high doses (Ames et al 1995). This suggests that there are a large number of chemicals in the environment presenting themselves as health hazards and the only acceptable method for testing is the rodent cancer bioassay. To date, the EPA has endorsed no commercial prediction system as a replacement for legally mandated testing of a toxicity health endpoint, although predictions can be submitted for consideration in overall chemical evaluations (Richard 1998). The reliability of rodent carcinogenicity assays is usually ascertained by repeating experiments with the same substance under the same test conditions (Gottmann et al 2001). However, given the cost of $2 -3 million and the three to five years required for planning, testing, and subsequent data analysis of a single chemical in a lifetime rodent carcinogenicity bioassay, "initial decisions on whether to continue development of a chemical, to submit pre-manufacturing notice (PMN), or to require additional testing may be based largely on structure-activity relationship models and short-term assays" (Klaassen 2001).

The standard two-year rodent bioassay used to determine the carcinogenicity is usually carried out on the suspected chemical at the maximum tolerated dose (MTD). The MTD can be defined as an estimate of the highest dose of an agent administered to a test animal in a chronic study that will not modify longevity from effects other than carcinogenicity. Due to this, it is speculated that these carcinogenic chemicals, most of which stem from industry, are carcinogenic and/or mutagenic at low doses in humans. Without further relevant data, this has to be considered true for human risk assessment. Simply, without data on mechanism of action for a given chemical, the true risk of cancer at low dose is highly uncertain, even for rats or mice (Gaylor et al 1995). By manipulating the experiment, "more than 90% of all chemicals can induce some tumor in a rodent" (Furst

1991). Pitfalls encountered in bioassays result from "not specifying the exact agent under test and how it relates to human exposure, using inappropriate routes of administration unrelated to humans, administering high doses, or concluding that a cancer was induced without adequate histopathological description of the lesion" (Furst 1991). Although the focus of this study was not geared toward uncovering the problems of *in vivo* assays, such pitfalls must be taken into consideration before any conclusions based on the results of this study can be drawn.

2.5 Comparison of SAR and QSAR Models

SAR has been applied both in a qualitative way (e.g., as simple recognition of structural alerts), and in a quantitative way (e.g., QSAR) to build mathematical models linking the physical chemical or structural properties of the molecules to the toxicological endpoints (Benigni et al 2004). The EPA has historically relied upon SAR for screening new chemicals for adverse health effects under the Premanufacture Notification Review requirements of the Toxics Substances Control Act (TSCA) (Wagner et al 1995). The EPA defines structure-activity relationship (SAR) as the relationship of the molecular structure of a chemical with a physicochemical property, environmental fate attribute, and/or specific effect on human health or an environmental species.

An SAR analysis can be either qualitative (SAR) or quantitative (QSAR). An SAR study seeks to identify the essential features of chemical structure that are determinants of a biological activity from available structural and bioassay information (Richard 1995). This implicitly assumes a causal relationship between fundamental molecular properties and activity in the bioassay under study, even when details of the mechanism of action are unknown (Richard 1995). According to the EPA, the difference between the two lie in that

qualitative SAR predictions are based on a comparison of valid measured data from one or more analogs (i.e., structurally similar compounds) with the chemical of interest. For example, categorical terms such as “similarly toxic”, “less toxic”, or “more toxic” would be used in a qualitative SAR assessment for toxicity to humans or environmental species. Quantitative predictions, on the other hand, are usually in the form of a regression equation and would thus predict dose-response data as part of a QSAR assessment (EPA 2002). Both SAR approaches can be applied to mechanism-based studies.

2.6 Expert Systems

Computational models with good predictive capability can be generated and several of these computational tools utilize substructural moieties (i.e., fragments) for identifying active regions of the molecules (White et al 2003). However, to date, assessments for health endpoints have been based primarily on chemical analogy and expert judgment and not on commercial prediction systems (Wagner et al 1995). The phrase “expert system for predicting toxicity” is used in various ways in the literature. Henceforth, for the purposes of this thesis, an expert system (ES) for predicting toxicity is considered to be “any formalized system, not necessarily computer-based, which enables a user to obtain rational predictions about the toxicity of chemicals” (Dearden et al 2003).

Therefore, in these terms, an expert system is a computer program that “provides solutions to important problems similar to those obtained by human experts” (Benfenati et al 1997). There are two classes of expert systems: (1) automated approaches that rely on the use of statistics in determining correlations between chemical structure and biological effect (machine learning programs); and (2) knowledge-based systems that rely on a set of programmed rules distilled from available knowledge and human expert judgment (Richard

1998). These two categories of approaches differ in the ways that they represent, process, and generalize chemical-biological activity information. Some commercially available toxicity prediction programs are DEREK, CASE/MultiCASE, TOPKAT, and ONCOLOGIC.

The CASE/MultiCASE program falls under the first category. The MCASE (Klopman and Rosenkranz 1994) and TOPKAT (Enslein 1994) approaches rely upon statistical or automated algorithms for extracting SAR associations from existing data, with little or no prior application of expert judgment or organization of data according to mechanism or chemical class. The ONCOLOGIC (Woo et al 1995) and DEREK (Sanderson and Earnshaw 1991; Greene 1996) approaches, in contrast, are rule-based systems built almost entirely upon prior knowledge, heuristics, expert judgment, and chemical and biological mechanism considerations (Richard 1999). DEREK makes qualitative rather than quantitative predictions by looking for previously characterized structural alerts and their associated toxic activity (Benfenati et al 1997). ONCOLOGIC has a typical structure consisting of a large number of rules that are combined together in order to give specific answers to specific questions pertaining to carcinogenicity (Woo 1998).

It is important to note that each individual toxicity prediction must be judged based not only on the statistical performance of the model (e.g., overall model sensitivity, specificity, accuracy), but also on the strength of the mechanistic or molecular analogy argument supporting the prediction. EPA's revised cancer Risk Assessment Guidelines, for example, specifically encourages the use of SAR information provided that it strengthens an argument based on "biological plausibility" and "mode of action consistent with generally agreed upon principles and understanding of carcinogenicity" (Richard 1999). Some of the benefits of an expert system are summarized as follow:

(i) Captures expertise before it is lost via use of existing data derived from human knowledge (e.g., DEREK and TOPKAT)

(ii) Reduces dependence upon one expert

(iii) Reduces errors and inconsistencies

(v) Increases knowledge-sharing

(vi) Expedites decision-making and may assist in the formulation of policies

Toxicity prediction systems are much more capable of, and useful for, identifying a potential toxicity hazard than they are at ruling out a hazard (Richard 1998). SAR models, in general focus primarily on determining the mechanistically bounded requirements for an activity, (i.e., identifying structural alerting features for activity). Lack of activity, representing a fundamentally unbounded condition, is most often predicted by the absence of activating features (Richard 1998). Although the methodology prescribed with expert systems involve computers, human expertise must be incorporated into the modeling process. In other words, good data is essential for SAR modeling and human experts are needed to define what data is actually good data. Also, once the computer-generated model makes predictions, a human expert familiar with the biology of the endpoint being tested is needed to interpret the data and draw acceptable conclusions. It is important to remember that SAR models are based upon toxicological testing and the ultimate goal is not always to replace animal testing, but to support the data and use mechanistic-based SAR modeling as an aiding tool to help in the prioritization of the vast amount of untested chemicals.

2.7 SAR Models as Classifiers

As of the last decade, SAR modeling has emanated as a useful tool in the screening of large, complex and noncongeneric (dissimilar) compounds. The historical success of its use

along with the experimental reproducibility of the rodent bioassay results in toxicological studies has triggered further research in the area. Hence, in addition to providing mechanistic data that's explainable by current knowledge, predictive SAR models present hypothetical cases. The cat-SAR approach as well as other SAR approaches, have been developed by ruminative evaluation of existing experimental rodent carcinogenicity results. For example, the models in this study were based on the toxicological results outlined by Gold and others in the Carcinogenic Potency Data Base (CPDB) (Gold et al 2001).

SAR models attempt to classify groups of chemicals based on their structural features that may or may not be associated with one or more biological activities. Hence, SAR datasets must be evenly balanced (Linusson et al 2000). If it is not, the applied learning set may consist of unbalanced data and a SAR model may be developed that has high sensitivity and low specificity yielding a high number of false positives or low sensitivity and high specificity yielding a high number of false negatives (Walker et al 2003). The goal is to have a high rate of prediction. However, risk attitudes govern the use of toxicity prediction systems for hazard identification and screening (Richard 1998). In assessing a toxicity hazard, a false negative prediction is much more costly from a risk management standpoint than a false positive prediction. For instance, if a chemical is determined to be safe and is allowed on the market there can be ecological and human consequences, whereas, a false positive prediction cost will be primarily in revenue for shareholders of a corporation if they had intended to market the product.

In respect to the cat-SAR approach, prediction accuracies are assessed on the basis of overall concordances with the CPDB carcinogenicity classifications of positive (carcinogenic) or negative (noncarcinogenic). For instance, the reproducibility of the CPDB

and *Salmonella* mutagenicity data are 75% (Gold 1991) and 85% (Zeiger 1985), respectively. Hence, these accuracies can be considered as estimates of the top accuracy reachable by any SAR method studying the toxicological phenomena. Most SAR approaches to carcinogenicity predictions share a clear limitation. That limitation is that they produce high sensitivities and low specificities. In other words, the prediction system incorrectly predicted many non-carcinogens as positive. It is believed that this may be due to the presence unbalanced datasets. The cat-SAR program, however, addresses this limitation by employing an optimization scheme.

2.8 Influence of Structural Diversity of Datasets on Model Prediction

The greatest challenge faced by users and designers of SAR programs is being able to predict carcinogenicity for a wide diversity of molecular structures, spanning an undetermined number of chemical classes and biological mechanisms (Richard et al, 2002). Whereas a drug design study deals with a well-defined class of congeneric chemicals that induce a well-defined biological activity, in toxicology one often deals with a certain biological effect (e.g., mutation, cancer) that can be induced by chemicals very different from each other (i.e., noncongeneric chemicals) (Benigni et al 1994). This recognition of the inherent biological and chemical compartmentalization of the carcinogenicity prediction problem as consisting of a large number of mechanistically distinct or overlapping sub-problems, is necessary for formulating effective solutions to the general prediction problem, as well as for understanding the limitations and successes of current approaches (Richard 1999).

The study presented herein is an attempt to determine the relationship between environmental chemicals and mammary carcinogens and evaluate the robustness and

predictability of the newly developed cat-SAR program. The evaluation of the SAR approach employed herein serves as a valuable illustration of its application for toxicity prediction. The results will be used to draw conclusions about the link between several end-points in terms of structural make-up. Furthermore, the derivation of final suitably validated cat-SAR models in this area may have significant utility as a two-dimensional data base searching tool, to the extent that large data bases of candidate molecules can now be rapidly screened and prioritized for compounds that should undergo subsequent toxicological evaluation as breast carcinogens. It will be seen from the findings in this study that the conclusions regarding mechanistic relatedness among toxicological phenomena can be successfully drawn with the use of cat-SAR.

CHAPTER 3. MATERIALS AND METHODS

3.1 Cat-SAR Study: Introduction

As mentioned, the categorical-SAR (cat-SAR) methodology is a new *in silico* approach to modeling toxicological phenomena. The program serves as a powerful analysis tool that can identify significant structural features that may be responsible for the experimental activity (i.e., mammary carcinogenicity) and inactivity (i.e., non-mammary carcinogenicity) of test compounds for a specific endpoint. For the purposes of this study, the toxicological endpoints evaluated were breast-specific and rodent carcinogens. The fundamental concept of the program is that if a substructure correlates with the experimental active chemical, it will be present primarily in active compounds in the dataset. On the other hand, if it is not related to the experimental activity, it will be randomly dispersed amongst active and inactive compounds of the database.

The program's capability in determining a compound's potential for being a carcinogen was assessed through the leave-one-out cross-validation (LOO-CV) procedure. LOO-CV is a re-sampling method used to assess the predictivity and stability of SAR models where one test compound is removed and the remaining compounds are used to predict its activity. Basically, the program involves a series of interrelated steps that when combined generate objective and unbiased models. The cat-SAR evaluation occurs in the following order: (1) database construction and molecular modeling, (2) *in silico* fragmentation of the learning set, (3) LOO-CV, (4) prediction, and (5) mechanistic analysis. Succinctly, the process begins by gathering a hierarchy of information by simply separating and classifying the learning set in question into relevant or purposeful chemical categories (e.g., carcinogens and noncarcinogens). The second step involves the entry of binary response data in terms of a

particular compound's biological activity (e.g., "1" corresponds to positive and "0" to negative compounds). The assigned value for classes of test compounds can be any whole number with the goal of making a clear distinction between different classes of test compounds. It should be noted that these values are not representative of the chemical's potency or toxicity level.

Foremost, this method is an attempt to enhance the general understanding of chemical-induced etiology of breast cancer. As previously mentioned, the main goal behind the proposed methodology is basically two-fold: (i) to investigate the potential that environmental chemicals may be involved in the development of breast cancer and (ii) to evaluate the cat-SAR algorithm. It was speculated that the approach employed in this study would provide a clear distinction between mammary and non-mammary gland carcinogens and noncarcinogens.

3.2 Methodology

3.2.1 Carcinogenic Potency Database (CPDB)

Toxicological data were obtained from the CPDB compiled by Gold et al. (Gold et al 2001). The CPDB analyzes and consolidates into a single resource the world's diverse literature on chronic, long-term animal cancer bioassays (Gold et al 1995). Analyses of 6073 experiments on 1458 chemicals (i.e., positive or negative for carcinogenicity), that have been reported in Technical Reports of the National Cancer Institute/National Toxicology Program (NCI/NTP) or in papers in the general published literature are presented (Gold et al 2001). Therefore the CPDB consists of two different subsets, the results from the NCI/NTP and the results from the general literature. The CPDB plot standardizes the experimental results (whether positive or negative for carcinogenicity), including qualitative data on strain, sex, route of compound administration, target organ, histopathology, and the author's opinion and

reference to the published paper, as well as quantitative data on carcinogenic potency, statistical significance, tumor incidence, dose-response curve shape, length of experiment, duration of dosing, and dose rate (Gold et al 1999). This systematic route was taken to better investigate the relationship between the identified structural alerts and its carcinogenic activity.

Furthermore, a potency value for carcinogens, the TD_{50} is calculated. Gold and collaborators, in a simplified way, defined a TD_{50} as that dose rate in mg/kg body wt/day, which, if administered chronically for the standard life span of the species, will halve the probability of remaining tumorless throughout that period (Gold et al 1999). Stated differently, a TD_{50} is that daily toxic dose that is required to induce tumors in 50% of the test animals. For this approach, the categorizing of potency values involved either a "0" or a "1" where a 1 was simply assigned to those chemicals reported as carcinogens (i.e., assigned a TD_{50} value in the CPDB) and a 0 to those designated as noncarcinogens (i.e., chemicals without a TD_{50} value). This is termed a categorical approach to modeling.

Using this data, a general rat (946 chemicals), mouse (763 chemicals), rat female (723 chemicals), mouse female (738 chemicals), and several sets of rat (100 mammary carcinogens) and mouse (24 mammary carcinogens) mammary gland carcinogen models were constructed. Each of these models contained an "active" and "inactive" category. . In summary, the following mammary databases were created: 1) a model of mammary carcinogens (MC) categorized as the "active chemicals" and non-mammary carcinogens (NMC) as the "inactive chemicals". Hence, referred to as the "MC-NMC" model and 2) a model comprised of MCs and whole animal noncarcinogens (NC). However, for the mouse dataset, instead of using a random set of mouse noncarcinogens, a random set of rodent

noncarcinogens were used. Thus, this model was termed the "MC-NC" model for the rat and "MC-RNC" for the mouse. The reason for considering two distinctive "inactive" categories (i.e., NC and NMC categories) for the rat and mouse mammary carcinogen models is to assess organ-specific carcinogenesis (i.e., breast carcinogenesis) by distinguishing between the activating and inactivating fragments of breast carcinogens from that of noncarcinogens and also from carcinogens that do not induce breast carcinogenesis. This approach aids in demonstrating that the set of mammary carcinogens in question are in fact inducing tumors in the mammary gland and not just a chance occurrence of predicting mammary cancer.

Moreover, to eliminate the possibility of attaining a "true model" via chance occurrence, models for each fragment set of randomly selected "inactive" compounds, albeit noncarcinogens or non-mammary carcinogens were constructed in multiples of three. In other words, three models were produced for both the MC-NMC and MC-NC models. With respect to the rat MC-NMC and MC-NC models, each model contained the same 100 rat mammary carcinogens (MC), as outlined in the CPDB (Gold et al 2001), representative of the model's "active" category. However, three of these models contained 100 randomly selected rat noncarcinogens (NC) as the "inactive" category. The other three models were comprised of 100 randomly selected non-mammary gland carcinogens (NMC) (i.e., general rat carcinogens) for its "inactive" category.

Likewise, of the 48 chemicals that comprised each of the three mouse MC-RNC and MC-NMC learning sets, 50% were listed in the CPDB as mouse mammary gland carcinogens (i.e., the model's "active" category). Hence, each of the six mouse mammary carcinogen models contained these 24 MCs as the model's "active" category. On the other hand, three of

the models constructed contained 24 randomly selected rodent noncarcinogens (RNC) as the "inactive" category as previously described.

3.2.1.1 Selection of Mammary Carcinogens

The mammary carcinogen models are based on all mammary carcinogens in the CPDB. Overall, the CPDB mouse and rat mammary gland consisted of 24 and 102 chemicals, respectively. Of the 102 rat mammary gland carcinogens (including several male-only breast carcinogens), 100 were suitable for our present analysis. Due to program restrictions on complex polymers, racemic mixtures, technical-grade chemicals, metals, metaloorganiccompounds, and complex inorganic compounds, norlestrin (a complex mixture) and dimethylaminoethylnitrosoethylurea, nitrite salt were excluded from the learning set. Moreover, organic salts were included as the freebase. Simple and defined mixtures and technical grade compounds were included as their major or active component.

All of the 24 CPDB mouse breast carcinogens met our set criteria and hence, all were included in the database. An equivalent number of non-carcinogens were then randomly obtained and added to the rat and mouse databases giving rise to a total of 48 mouse chemicals (50% active and 50% inactive compounds) and 200 rat chemicals (50% active and 50% inactive compounds).

3.2.1.2 Selection of Mammary Noncarcinogens

The mouse noncarcinogens (i.e., "inactive" category) were randomly selected from a database of 47 chemicals tested to be inactive (no observed tumor induction) in rodents. However, due to the small number of chemicals that tested negative in both the rat and mouse species, we were unable to mesh the 100 rat mammary gland carcinogens with an equivalent number of rodent noncarcinogens. So instead, a total of 100 rat noncarcinogens were randomly selected from a set of 449 CPDB chemicals tested negative in only the rat

species. Likewise, for the rat mammary-specific model, a set of 100 compounds was randomly selected from a total of 395 (excluding the 100 mammary carcinogens) CPDB carcinogens all of which did not induce mammary tumors in the rat species.

Furthermore, to assure random assortment of the "inactive" category of the models, three sets of random selection was performed for each rat and mouse mammary-carcinogen model constructed. It should be noted that each inactive chemical in the learning set had an equal opportunity of being selected. For example, after selecting a group of chemicals from the dataset to represent the "inactive" category of the model, these chemicals were placed back into the pool of noncarcinogens for a second draw. This was repeated until all three models were comprised of an equal set of noncarcinogens. The rationale behind this approach was to ensure model consistency among random inactive subsets of the database. Hence, each of the models constructed is an illustration of the "true model" which demonstrates how stable (or consistent) the individual model estimates are.

This multiple random sampling also helps in determining the dependency of the model on the dataset on which it was built. This procedure helps: 1) eliminate to some extent the representation bias in the database due to its composition and random selection, 2) determine the variability and stability of the resulting SAR models derived from random samplings of the empirical distribution, 3) identify the most consistent model, and 4) determine the overall reliability of the "true model" on the dataset from which it was built.

3.2.1.3 General Rat and Mouse Learning Sets

Two databases containing all tested rat and mouse chemicals (actives and inactives) that met our set criteria were also developed, fragmented into all possible fragments, and underwent validation and prediction by the cat-SAR program. Again, data used to build these

models were entirely based on information gathered by Gold and others in the year 2001 plot of the CPDB (Gold et al 2001). The rat model contained 946 compounds in which 447 were inactive compounds and 499 actives. The mouse model contained 769 chemicals in which 386 inactive chemicals and 383 active chemicals. It was then expected that a greater number of structural alerts would be generated due to the size of these learning sets. The rodent mammary gland models discussed earlier are subsets of these general models. Stated differently, the compounds analyzed in the rat mammary gland models were also considered in the general rat model. The same is true for those carcinogens and noncarcinogens contained in the female mouse mammary gland models.

3.2.2 Database Construction

Just like all SAR approaches the cat-SAR approach begins with a given set of chemical compounds with their known structures and associated biological activity. The cat-SAR datasets are composed of its chemical name, molecular structure, and categorical label (i.e., "0" or "1"). Tripos® SYBYL6.8 Holographic Quantitative Structure-Activity Relationship (HQSAR©) molecular modeling software package (Tripos Associates, Inc., St. Louis, Missouri) was used for the *in silico* fragmenting of all learning sets of chemicals into all possible fragments whereas cat-SAR was used to do all SAR modeling and statistical analyses of compounds, validations, and predictions of chemical toxicity.

3.2.2.1 Significance of Balanced cat-SAR Datasets

For SAR studies, the chemical composition of learning sets is of critical importance. Most databases publicly available for the development of SAR models are not equally distributed (e.g., most often toxicologically-active chemicals pre-dominate) (Rosenkranz and Cunningham 2001). By having an equal number of inactive and active compounds in our models ensures that we can apply the "common-sense" or "weight-of-evidence" approach to

analysis. Additionally, a more accurate and optimal observed correct prediction (OCP) rate is expected. However, as seen with the general rat model this statement does not hold true. To avoid experiencing the problem of having a high presence of carcinogens predicted correctly and also a high number of noncarcinogens being predicted incorrectly due to a greater concentration of carcinogens versus noncarcinogens in the data set, the cat-SAR system can be adjusted to account for this indifference.

3.2.2.2 Model Parameters

Unlike most SAR approaches, the structural property of a fragment that goes into the final model is what drives the model to make a prediction and not the chemical compounds from which they originated. In cat-SAR, the ability to adjust fragment parameters (e.g., atom, bonds, connections, hydrogen atoms, chirality, and donor and acceptor) of the model can lead to an enhanced quality of the output data for the model. For this study, a hologram length of 151 was arbitrarily used for each of the models created and only the atoms (A), bonds (B), connections (C), and/or hydrogen (H) as a fourth parameter was selected. A fragment size of 3 to 7 heavy atoms (e.g. excluding hydrogen atoms) was selected for the derived models and compounds were designated as inactives (e.g., noncarcinogens with significantly inactive fragments) and actives (e.g., carcinogens with significantly active fragments). The various modeling parameters that can be adjusted by the user make the cat-SAR algorithm a more flexible program by providing greater opportunities to ascertain the relationship between chemical structure and toxicological activity thus separate it from existing commercial SAR programs.

The biological activity associated with each fragment was determined later in the study. This was done by examining the fragment parts of a chemical as it appears in the validation summary and based on weight-of-evidence we will be able to determine if a

chemical contains significantly more active than inactive fragments. From this observation, it was reasonable to conclude that a chemical was active versus inactive based on its significantly high presence of active molecular fragments (e.g., weight-of-evidence). This made it feasible to say that a particular chemical is indeed active or inactive based on its fragment composition. In other words, the selection of mechanistically important molecular descriptors can possibly reveal or relate to the structural basis for toxicological activity or inactivity.

3.2.3 *In silico* Fragmentation of Test Compounds

Upon completion of the initial fragment list for the learning sets, a Tripos SYBYL Add-on script was utilized for fragment counting. The Add-on script associates each fragment with all the molecular compounds in which it was found in the form of a tabulated compound-fragment matrix for each set of fragments (e.g., "ABC" and "ABCH" fragment sets). As previously stated, each fragment is labeled with the name and activity of its parent chemical. Upon completion of this process, cat-SAR organizes the list of fragments and tabulates the number of chemicals containing each of them (Figure 3.1). In other words, the generated fragments are examined by cat-SAR for its association with the toxicological endpoint in question. The fragment list will be used in investigating the mechanisms and also in predicting the activity of untested chemicals as well as compounds with unknown biological activity. The number of active and inactive compounds containing a specific fragment was then calculated.

Those chemical compounds containing "insignificant" fragments according to the set criteria will be excluded from the final model. Insignificant fragments are those structures within the initial model that are either: 1) unique, 2) found approximately equivalent in active and inactive compounds, and 3) did not meet the other user's criteria (discussed in next

subchapter). For example, the chemical fragment may be so unique that the fragment was only listed in one or two chemicals and thus, not predicted. However, chemicals that were found to contain "significant" active and inactive fragments or 'structural alerts' by the program will serve as the basis for the final model's make-up. Moreover, it should be noted that the selection of "significant" fragments utilizes the "Common Sense" approach and not statistical analysis.

3.2.4 Set Criteria: cat-SAR Rules

3.2.4.1 Selection of Significant Fragments

To establish a link between each molecular fragment and its associated activity and inactivity, a set of rules were derived for the selection of "significant" active and inactive fragments. The first selection rule entails the frequency or occurrence of the fragment's appearance in the dataset under consideration. This was conceptually set at three chemicals for all learning sets under evaluation. The primary reason for setting the cut-off point at three is to possibly 'rule out' or exclude those unique fragments that may not serve as being useful in terms of determining the overall structure-activity relationship. Based on this cut-off, in order for the program to call a given fragment a "significant" active or inactive (i.e., fragments used to make predictions), that fragment must first be identified in a minimum of three chemicals in the learning set. Hence, the exclusion of rare fragments makes the model more mechanistically informative.

To assess the phenomenon of possible exclusion of active and inactive structural alerts, it was presumed that any requirement less, or due to the small datasets, greater than three, would require fragments found by chance to be included in the model. This restricts the probability of losing rather important information (e.g., certain structural features) about a particular model. In other words, it was presumed that if the models were made too

restrictive taking into consideration the small size of the learning set (e.g., requiring a fragment to be found in more than three compounds), then vital information about the model could very well be lost in the process. However, it is noteworthy to mention that all fragments that fall outside this criterion are not necessarily "insignificant". This concept should not be over-looked as the number of cases in which those fragments found in one or two of the test compounds was simply inadequate in number and thus, not used by the program. The possibility of having "significant" fragments excluded from the final model based on this rule will be further analyzed.

The second selection rule involved the learning set's distribution of active and inactive compounds linked to a particular fragment. To fulfill the requirements of this rule, two proportions, 0.75 and 0.90 models, were developed for both the ABC and ABCH fragment sets. For instance, at a cut-off of 75%, if a fragment were present in twenty chemicals in the learning set it would have to be found in at least fifteen active or inactive chemicals to be considered a significant fragment. However, when set at 0.90, the selection process becomes more stringent as a fragment must now be found in at least ninety percent of the total number of active or inactive compounds in the learning set, and also meet the requirement of the first rule in order for it to be identified as either a "significant" active or inactive chemical. In other words, this criterion in comparison to the 75% proportion imposes a stricter requirement when selecting the "significant" fragments from the total pool of fragments.

It is important to note that the very same fragments are being used for both the 0.75 and 0.90 proportions. Hence, all the generated fragments listed in the 0.90 proportions will also be found in the 0.75 proportions. Once again, the user selects these proportions. If the

criteria set by the user are not met, those fragments are assumed to be "insignificant" and will be excluded from the final model (composed of only "significant" active and inactive fragments). As a result of this screening process, a much smaller fragment set is generated and will go into constructing the final model.

Moreover, it was reasoned that if a particular fragment is associated with activity, there may yet be other reasons (i.e., fragments) for it being inactive, thus it would not be expected to be found in 100% of the active compounds. Likewise is true for inactive fragments. Thus, if consideration was based largely on fragments found exclusively in active or inactive compounds, the fragment pool is rarified to an unacceptable level and there is an associated risk of the model becoming less information-intensive (i.e., loss of important information). Of importance, fragments found equally in active and inactive compounds are also eliminated. Such fragments may serve as chemical scaffolds holding the biologically active features and are not directly related to activity or inactivity (Cunningham et al; in press 2005).

In summary, fragments were considered "significant" if they were found in at least three compounds in the dataset and depending on the model, also found in at least 75% or 90% of the active or inactive compounds that derived them. The ensuing "significant" fragments are used in predicting the toxicity of a test compound and also for mechanistic analysis. The models developed are listed in Tables 1-8.

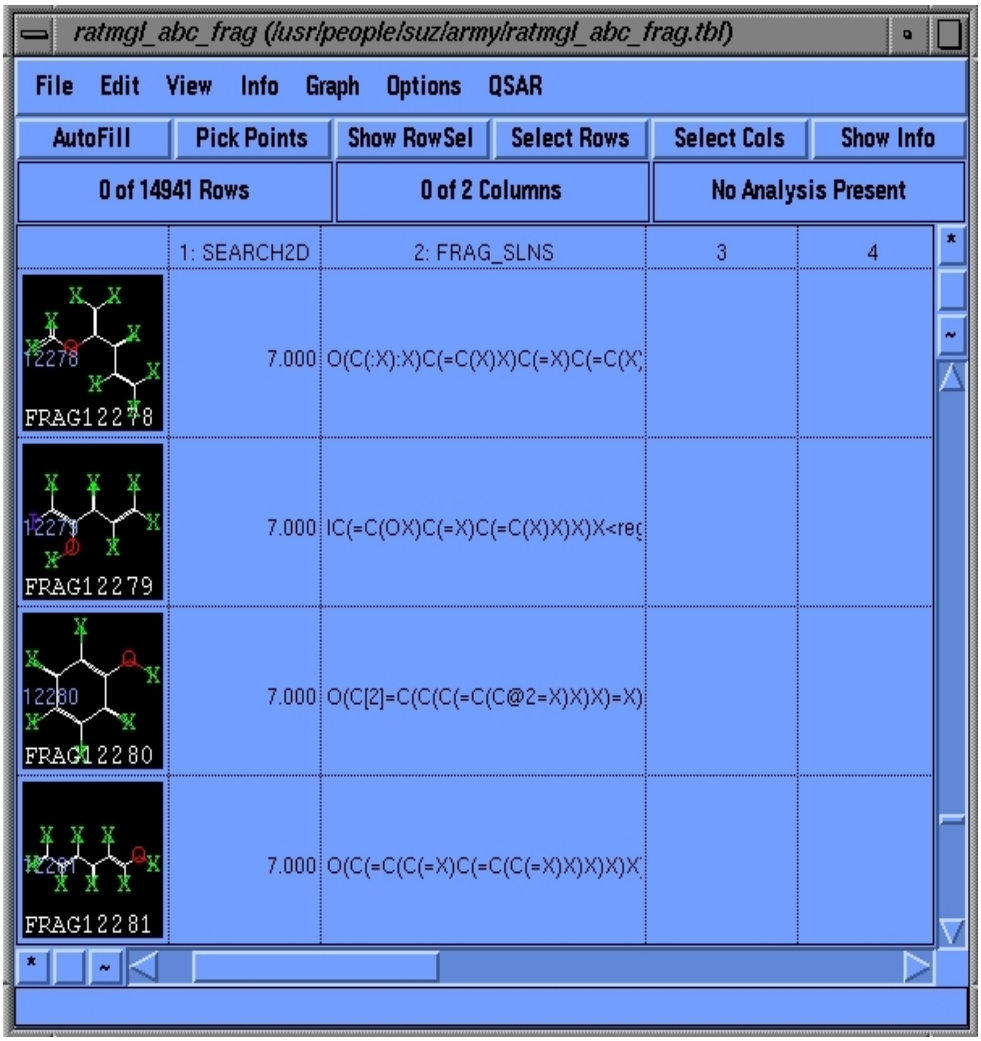

Figure 3.1 Tripos Sybyl HQSAR fragment-embedded molecular spreadsheet illustrating fragments 12,278 through 12,281 out of the 14,941 total generated fragments with coded atoms from the rat MC-NC model.

3.2.5 Model Predictions

The cat-SAR model determines if the molecule in question contains an active or inactive fragment from the model's dataset. The program does this by operating on a "weight-of-evidence" where a chemical is screened for fragments listed in the significant fragment list and whether or not it is categorized as an active or inactive depends on its structural composition. Briefly, if more fragments are found that are inactive versus active, then this compound will be predicted as inactive and vice versa. For example, if a compound

contains two fragments, one based on 5/6 inactives (i.e., fragment was found in 5/6 inactive compounds) and the other fragment was found in 3/3 active compounds in the model's learning set, then this unknown compound will be predicted by cat-SAR to have an 89% (i.e. 8/9) chance of being inactive for this endpoint. This is the probability of activity for that test compound. Likewise, the compound will have an 11% (i.e., 1/9) chance of being inactive. Hence, the probability of activity or inactivity can be computed on the basis of a test compound's fragment composition. In other words, this approach allows the probability of activity and inactivity to be determined by comparison of the structure of the test chemical against the entire structural information present in the model. Furthermore, if the model is unable to link any of the "significant" active or inactive fragments to a test compound, then no prediction is made for that compound.

3.2.6 Model Validations

The validation of SAR models is important because it assesses the model's reliability (Walker et al 2003). Herein, the models created for each database were tested for robustness and reliability. In order to estimate each model's quality, the cat-SAR algorithm adopted the LOO procedure to cross-validation. In these validation procedures, each chemical was removed from the complete dataset of chemicals (n), one at a time, leaving (n-1) chemicals as the learning set. This probability of activity or inactivity for each chemical was recalculated using the (n-1) database as the learning set. In other words, an SAR model was constructed from the remaining 99% of the chemicals. At this point, all molecular fragments from the (*n-1)* set are used in calculating the probability of a chemical being active or inactive. The LOO-CV procedure was repeated on each test compound using the reduced fragment set (i.e., the group of fragments that met the set criteria) n times until each

compound had been tested. As previously mentioned, the probability of activity or inactivity determines the model's prediction of a test compound. Hence, it is possible to determine the overall probability of activity or inactivity to be used in categorizing a chemical as either a carcinogen or noncarcinogen. This enables the user, based on the results from the LOO-CV, to select and apply an optimal cut-off point that aids in the separation of predicted active and inactive compounds. Validations were performed for each database. The cat-SAR approach is applicable to various learning sets of dissimilar compounds.

3.2.6.1 Statistical Evaluation: Sensitivity, Specificity, and OCP

The two concerns addressed when studying the cat-SAR models were: (1) the predictive capability of the model, and (2) the consistency of the model form. The success of a predictive SAR model can be described in a combination of ways. In general, the validity of SAR models based on learning sets is assessed by the sensitivity, specificity, and concordance (i.e., observed correct predictions (OCP)) statistics when describing phenomena.

Simply put, the sensitivity is defined as the number of correct positive predictions out of the total number of positive predictions made. The sensitivity of a model expresses the ability of the model to accurately predict a true positive chemical as positive, i.e., in the present carcinogenicity study, carcinogenic (Walker et al 2003). In contrast, specificity is the total number of correct negative predictions out of the total number of negative predictions made. The specificity of a model expresses the ability of the model to accurately predict a true negative chemical as negative, i.e., in the present study, noncarcinogenic (Walker et al 2003).

On the other hand, the concordance (i.e., OCP rate) is defined as the number of correct predictions (i.e., correctly predicted active and inactive test compounds) out of the

total number of predictions made. Generally the results of this study summarized concordant and discordant classifications (carcinogen and noncarcinogen) attained by the cat-SAR program. The sensitivity and specificity of the SAR models developed herein were determined by the removal of a test compound from the database.

3.2.7 Mechanistic Rationalization: Chemical Diversity Approach

While the objective of this study attempts to support the widely studied hypothesis that models designed *in silico* are as valuable as animal bioassays, it was necessary to determine, if any, mechanistic overlap occurring between models of different toxicological endpoints (i.e., *Salmonella* mutagenicity, carcinogenicity, mammary-specific carcinogenicity, and estrogenicity). Furthermore, while reliable databases of toxicological phenomena, when available, are usually small in chemical content, the approach used herein predicts the toxicological profiles of 10,000 chemicals (Rosenkranz et al 2000). The "Chemical Diversity Approach" (CDA) operates on the assertion that the mechanistic relationship between toxicological phenomena can be determined on the basis of the prevalence of chemicals tested in an assay to have the same biological activity. These chemicals were derived from chemical structure libraries and from a random sample of molecular structures from the National Cancer Institute Repository of potential cancer chemotherapeutic agents (Cunningham et al 2004). By employing the CDA, it was possible to analyze the potential of chemicals demonstrating breast carcinogenicity in the standard rodent assay to be possibly related to other toxicological phenomena such as mutagenicity and estrogenicity.

The applicability of this approach is that while no SAR model is perfectly predictive, when applied to a population of 10,000 chemicals, provided the sensitivity and specificity are approximately equal we can expect that the overall prevalence will reflect the true distribution (Rosenkranz et al 2000). This allows a determination of the significance of the

observed joint prevalences. Under the assumption (i.e., the null hypothesis), that there is no relationship between the two or more toxicological phenomena under investigation, the observed joint prevalence of chemicals that induce both phenomena could then be compared with the expected joint prevalence. If the observed prevalence is significantly greater than the expected one, then it can be concluded that the two phenomena are related to one another mechanistically. Similarly, if the observed prevalence is significantly lower than the expected one, it suggests that the two phenomena are antagonistic with one another, e.g., they could compete for an active site.

In implementing such an approach, mechanistic insight into the breast carcinogenicity phenomenon is gained by evaluating the concordance, or lack thereof, between biological endpoints. For instance, the CDA is applicable when considering the mechanistic relationship between agents that are *Salmonella* mutagens and those that cause the breast carcinogenicity phenomena (i.e., mitogenesis and binding to the estrogen receptor). Additionally, the approach can be used to confirm specific hypotheses (e.g., the electrophilic theory of cancer causation) as well as to generate new (i.e., knowledge-based) hypotheses driven solely by the data and the availability of appropriate SAR models (Rosenkranz et al 2000).

3.2.8 Overview of cat-SAR Method

In review, two toxicological phenomena were studied. The first phenomenon was cancer and the other was mammary-specific carcinogenesis. All data was taken from the 2001 version of the CPDB (Gold et al 2001) and all statistical analysis was performed with the cat-SAR algorithm (see Tables 1-8). For each learning set, an ABC (atoms, bonds, and connections) and ABCH (atoms, bonds, connections, and hydrogen) sets of fragments were created. These sets are important in fragment distinction. Furthermore, each generated model

consisted of an “active” and “inactive” category. The varying predictive performance of these groups assisted in identifying an optimal model. The CDA is then applied to the resulting optimal models and a group of 10,000 noncongeneric (i.e., dissimilar) chemicals to determine mechanistic relatedness between different toxicological endpoints.

CHAPTER 4. RESULTS AND DISCUSSION

4.1 Study Results

4.1.1 Predictability of cat-SAR Mammary Carcinogen Models

The models assessed for predictivity were the rat and mouse ABC and ABCH mammary carcinogen-non-mammary carcinogen (MC-NMC) and the mammary carcinogen-non-carcinogen (MC-NC or MC-RNC) models. During LOO-CV, sensitivity, specificity, and concordance values of the models were established. Through LOO-CV, the rat MC-NMC ABC 90% model presented the best predictivity. This model attained a sensitivity of 83% and a specificity of 74% yielding a concordance of 79% (ABC 3/0.90 Model 1, Table 4.1). Predictions were made on 124 of the chemicals in the learning set (Table 4.5).

Table 4.1 Predictive performance summary for the rat mammary–non-mammary carcinogen (MC-NMC) cat-SAR model. The ABC model was based on fragments of size between three and seven heavy atoms and considered atoms, bonds, and atom connections. The ABCH model included consideration of hydrogen atoms.

Model (opt. 0.51)	Total Fragments	Model Fragments	Active Fragments	Inactive Fragments	Sensitivity	Specificity	OCP
ABC 3/0.75							
Model 1	13868	1349	849	500	0.80(70/88)	0.66(53/80)	0.73(123/168)
Model 2	14461	1330	861	469	0.72(63/87)	0.72(59/82)	0.72(122/169)
Model 3	14427	1245	767	478	0.68(59/87)	0.74(64/86)	0.71(123/173)
Average	14252	1308	825	482	0.73(64/87)	0.71(59/83)	0.72(123/170)
ABC 3/0.90							
Model 1	**13868**	**1102**	**731**	**371**	**0.83(58/70)**	**0.74(40/54)**	**0.79(98/124)**
Model 2	14461	1086	723	363	0.82(54/66)	0.69(44/64)	0.75(98/130)
Model 3	14427	847	520	327	0.82(51/62)	0.72(41/57)	0.77(92/119)
Average	14252	1012	658	354	0.82(54/66)	0.71(42/59)	0.77(96/124)
ABCH 3/0.75							
Model 1	32235	3679	2081	1598	0.81(78/96)	0.62(55/89)	0.72(133/185)
Model 2	32374	3921	2088	1833	0.70(66/94)	0.64(59/92)	0.67(125/186)
Model 3	32627	3497	1928	1569	0.75(70/93)	0.69(65/94)	0.72(135/187)
Average	32412	3699	2032	1167	0.76(71/94)	0.65(60/92)	0.70(131/186)
ABCH 3/0.90							
Model 1	32235	2750	1642	1108	0.81(65/80)	0.76(50/66)	0.79(115/146)
Model 2	32374	2947	1637	1310	0.75(55/73)	0.69(53/77)	0.72(108/150)
Model 3	32627	2241	1170	1071	0.81(63/78)	0.70(52/74)	0.76(115/152)
Average	32340	2646	1483	1163	0.79(61/77)	0.72(52/72)	0.76(113/149)

Footnotes

(table cont.)

Total Fragments: number of fragments derived from learning set.
Model Fragments: number of fragments meeting specified rules of the model.
Active Fragments: number of fragments meeting specified rules to be considered as active.
Inactive Fragments: number of fragments meeting specified rules to be considered as inactive.
Sensitivity: number of correct positive predictions/total number of positives predicted.
Specificity: number of correct negative predictions/total number of negatives predicted.
Observed Correct Predictions: number of correct predictions/total number of predictions.

Likewise, the best rat MC-NC model with the best predictive performance was the ABC 90% model. This model was able to achieve a concordance between experimental and predicted results of 82% with a sensitivity of 77% and a specificity of 88% (ABC 3/0.90 Model 2, Table 4.2). For this model, predictions were made on 145 of the 200 chemicals in the dataset (Table 4.6). As noted, both rat MC-NC and MC-NMC models favored the stricter requirement (i.e., a proportion of 90%).

Table 4.2 Predictive performance summary for the rat mammary carcinogen – noncarcinogen (MC-NC) cat-SAR model. The ABC model was based on fragments of size between three and seven heavy atoms and considered atoms, bonds, and atom connection. The ABCH model included consideration of hydrogen atoms.

Model (opt. 0.37)	Total Fragments	Model Fragments	Active Fragments	Inactive Fragments	Sensitivity	Specificity	OCP
ABC 3/0.75							
Model 1	18021	1336	758	578	0.73(66/90)	0.78(69/88)	0.76(135/178)
Model 2	17369	1486	786	700	0.71(67/95)	0.80(72/90)	0.75(139/185)
Model 3	15547	1629	737	892	0.69(62/91)	0.76(67/88)	0.72(129/179)
Average	16979	1484	760	723	0.71(65/92)	0.78(69/89)	0.74(134/181)
ABC 3/0.90							
Model 1	18021	1016	642	374	0.82(62/76)	0.78(47/60)	0.80(109/136)
Model 2	**17369**	**1129**	**617**	**512**	**0.77(56/73)**	**0.88(63/72)**	**0.82(119/145)**
Model 3	15547	1311	624	687	0.83(63/76)	0.73(44/60)	0.79(107/136)
Average	16979	1152	628	524	0.80(60/75)	0.80(51/64)	0.81(112/139)
ABCH 3/0.75							
Model 1	38797	3859	1790	2069	0.72(68/94)	0.76(68/90)	0.74(136/184)
Model 2	37636	4293	2007	2286	0.71(70/98)	0.77(75/97)	0.74(145/195)
Model 3	34407	4093	1785	2308	0.73(71/97)	0.65(62/95)	0.69(133/192)
Average	36947	4082	1861	2221	0.72(70/96)	0.73(68/94)	0.73(138/190)
ABCH 3/0.90							
Model 1	38797	2746	1434	1312	0.76(63/83)	0.78(61/78)	0.77(124/161)
Model 2	37636	2923	1392	1531	0.75(63/84)	0.78(67/86)	0.77(130/170)
Model 3	34407	2949	1372	1577	0.74(66/89)	0.71(52/73)	0.73(118/162)
Average	36947	2873	1399	1473	0.75(64/85)	0.76(60/79)	0.76(124/164)

Footnotes: See Table 4.1

The best mouse MC-NMC model selected for further analysis was the ABC 75% model. This model yielded a concordance of 81% with a sensitivity of 80% and a specificity of 82% (ABC 3/0.75 Model 2, Table 4.3). Predictions for this model were made on 42 of the 48 chemicals comprising the dataset (Table 4.7).

Table 4.3 Predictive performance summary for the mouse mammary carcinogen–non-mammary carcinogen (MC-NMC) cat-SAR model with 3 to 7 heavy atoms.

Model (opt. 0.76)	Total Fragments	Model Fragments	Active Fragments	Inactive Fragments	Sensitivity	Specificity	OCP
ABC 3/0.75							
Model 1	5553	188	136	52	0.75(15/20)	0.61(11/18)	0.68(26/38)
Model 2	**4718**	**138**	**69**	**69**	**0.80(16/20)**	**0.82(18/22)**	**0.81(34/42)**
Model 3	6508	169	87	82	0.75(15/20)	0.78(14/18)	0.76(29/38)
Average	5593	165	97	68	0.75(15/20)	0.74(14/19)	0.77(30/39)
ABC 3/0.90							
Model 1	5553	106	73	33	0.80(12/15)	0.50(4/8)	0.70(16/23)
Model 2	4718	116	62	54	0.79(15/19)	0.78(7/9)	0.79(22/28)
Model 3	6508	122	69	53	0.83(15/18)	0.67(4/6)	0.79(19/24)
Average	5593	115	68	47	0.82(14/17)	0.63(5/8)	0.76(19/25)
ABCH 3/0.75							
Model 1	13517	801	591	210	0.62(13/21)	0.78(18/23)	0.70(31/44)
Model 2	12040	655	386	269	0.81(17/21)	0.77(17/22)	0.79(34/43)
Model 3	15187	753	434	319	0.62(13/21)	0.91(21/23)	0.77(34/44)
Average	13581	736	470	266	0.67(14/21)	0.83(19/23)	0.75(33/44)
ABCH 3/0.90							
Model 1	13517	443	324	119	0.55(11/20)	0.55(6/11)	0.55(17/31)
Model 2	12040	544	329	215	0.84(16/19)	0.74(14/19)	0.79(30/38)
Model 3	15187	553	352	201	0.79(15/19)	0.63(5/8)	0.74(20/27)
Average	13581	513	335	178	0.74(14/19)	0.62(8/13)	0.69(22/32)

Footnotes: See Table 4.1

In addition, the mouse MC-RNC model was the most predictive model and achieved a concordance of 81% with a sensitivity of 84% and a specificity of 76% (ABC 3/0.75 Model 1, Table 4.4). This model made predictions on 36 of the 48 chemicals in the learning set (Table 4.8). In contrast to the rat MC models, these models favored the least restrictive requirement (i.e., a proportion of 75%). Overall, all models achieved concordances significantly better (χ^2 =7.98, p=0.005) than chance (i.e., 50%). In fact, each of the mammary carcinogen models presented good predictive power indicative that the database's structural

alerts are artifacts of chance-correlation, but are relevant to the specific activity or inactivity categories they represent. In addition, almost all chemicals from the learning sets received a prediction (i.e., wide coverage). This helps substantiate the validity of the cat-SAR expert system in developing meaningful and generalized models.

Table 4.4 Predictive performance summary for the mouse mammary carcinogen – rodent noncarcinogen (MC-RNC) cat-SAR model with 3 to 7 heavy atoms.

Model (opt. 0.30)	Total Fragments	Model Fragments	Active Fragments	Inactive Fragments	Sensitivity	Specificity	OCP
ABC 3/0.75							
Model 1	**6414**	**379**	**72**	**307**	**0.84(16/19)**	**0.76(13/17)**	**0.81(29/36)**
Model 2	6504	357	185	172	0.72(13/18)	0.65(11/17)	0.69(24/35)
Model 3	6157	294	172	122	0.75(12/16)	0.83(15/18)	0.79(27/34)
Average	6358	343	143	200	0.78(14/18)	0.76(13/17)	0.77(27/35)
ABC 3/0.90							
Model 1	6414	352	84	268	0.87(13/15)	0.44(4/9)	0.71(17/24)
Model 2	6504	244	192	52	0.86(12/14)	0.40(4/10)	0.67(16/24)
Model 3	6157	195	109	86	0.86(12/14)	0.75(6/8)	0.82(18/22)
Average	6358	264	128	135	0.86(12/14)	0.56(5/9)	0.74(17/23)
ABCH 3/0.75							
Model 1	14963	1396	436	960	0.85(17/20)	0.68(13/19)	0.77(30/39)
Model 2	15956	1502	672	830	0.63(12/19)	0.58(11/19)	0.61(23/38)
Model 3	14819	1188	658	530	0.79(15/19)	0.79(15/19)	0.79(30/38)
Average	15246	1362	589	773	0.79(15/19)	0.68(13/19)	0.74(28/38)
ABCH 3/0.90							
Model 1	14963	1346	466	880	0.72(13/18)	0.80(12/15)	0.76(25/33)
Model 2	15956	1022	634	388	0.77(13/17)	0.65(11/17)	0.71(24/34)
Model 3	14819	1010	607	403	0.77(13/17)	0.67(10/15)	0.72(23/32)
Average	15246	1126	569	557	0.76(13/17)	0.69(11/16)	0.73(24/33)

Footnotes: See Table 4.1

4.1.2 Analysis of Mammary Carcinogen Models

To address model consistency, statistical analysis was performed on each of the three random selections of noncarcinogens and non-mammary carcinogens (i.e., model's "inactive" category). As a reminder, all models, regardless of ABC and ABCH assortment or inactive category make-up, are subsidiaries of the general rat and mouse models. The random selection of "inactive" compounds for the rat and mouse mammary carcinogen datasets shows that the models (models 1-3) for each of the ABC and ABCH sets are performing

about the same. In other words, the models differing in only non-carcinogens are statistically consistent with the total number of correct chemical predictions that is being made (i.e., similar OCP rates). For example, the rat MC-NC ABC 90% models all achieved OCP rates ranging from 79-82% (Table 4.2). Similar observations were seen with the other mammary carcinogen models (Tables 4.1- 4.4). Model consistency among the random subsets of inactive models is important because it assures that the models are not arbitrarily making predictions or mechanistic assertions. Despite the challenge of small datasets as presented by the mouse mammary carcinogen model, the cat-SAR algorithm was able to make predictions on a great number of compounds (i.e., greater than chance). Moreover, the models were all statistically consistent.

4.1.2.1 ABC and ABCH 75% and 90% Models

When making comparisons between the ABC and ABCH 75% and 90% models, the average values for fragment count, OCP, sensitivity, and specificity values were considered. Based on this, it was noted that the 90% criteria for selecting important fragments from the rat mammary dataset presented models with better OCP, sensitivity, and specificity. However, it should be noted that a cost is associated with this increased accuracy. This cost is a limitation in the model's mechanistic capability due to its exclusion of important fragments. It was observed that fewer predictions were made on chemicals with this stringent criterion. For example, the rat MC-NC ABC 75% model 2 made predictions on 185 chemicals in the learning set whereas the rat MC-NC 90% model 2 made predictions on 145 chemicals (Tables 4.2 and 4.8). Thus, the model is comprised of fewer significant active and inactive fragments possibly making the model less-information intensive. It is speculated that the decrease in the number of significant fragments from 1486 to 1129 (ABC 3/0.75 and 0.90

Model 2, Table 4.2) may be responsible for the reduction in the total number of chemical predictions made.

In contrast, the much smaller mouse mammary learning set favored the 75% criteria. Initially, this was expected because small-sized models would be expected to generate the least number of fragments (i.e., less information). Thus, it was reasonable to assume that based solely on the small sample size (i.e., 48 compounds comprised the model), it was very unlikely that many of these fragments would have been found in 90% or more and in at least 3 of the total active or inactive compounds in the learning set. Although, the mouse MC-NC (ABC 90%) model 3 when compared to model 1 of the ABC 75% set, has a higher OCP rate (Table 4.4), when comparing the two models in terms of the number of chemicals it predicted, model 1 was the better model overall. Simply, model 3 was a poor model to illustrate as such few inactive compounds were predicted. Thus, the OCP rate in addition to the total number of chemicals predicted by the model was justifiably taken into consideration when determining which of the three random subsets of inactive models was better in terms of its predictive power.

Furthermore, the ABC models had a better predictive performance when compared to the ABCH models, which were more specific by the inclusion of hydrogen atoms. Generally, it was noted that going from the ABC to the ABCH model resulted in doubling of the total number of fragments. For example, the rat MC-NMC ABC 90% models had an average of 14252 fragments and the ABCH model was based on an average of 32340 fragments (ABC and ABCH 3/0.90, Table 4.1). Similar observations were seen with the model's active and inactive fragments and also the model's total generated significant fragments (Table 4.1).

However, despite the doubling of significant fragments and an increase in the number of predictions being made, the ABCH models generally did not result in better sensitivity, specificity, and OCP rates. In most cases, this "doubling effect" resulted in decreased OCP rates of the mammary carcinogen models. Only in few instances, the OCP rate was slightly improved when going from the ABC model to the ABCH model. For example, the mouse MC-NC ABC 3/0.90 model 2 achieved an OCP of 67% whereas its ABCH counterpart yielded an improved OCP of 71% (Table 4.4). However, based on the computed averages for the ABC and ABCH 75% and 90% models, the ABCH models did not contribute to the model's enhancement of the overall predictive performance.

Table 4.5 Model validation for rat mammary gland carcinogens and non-mammary gland carcinogens (MC-NMC). For the ABC model, compounds with values equal to or greater than 51% were predicted to be carcinogenic and those below 51% were predicted to be non-carcinogenic. For the ABCH model, compounds with values equal to or greater than 63% were predicted to be carcinogenic and those below 63% were predicted as non-carcinogenic.

Model 1				Model 3-7/0.90	
Chemical	*Salmonella* Mutagenicity	CASN	Experimental Activity	ABC %Active (0.51)	ABCH %Active (0.63)
1-Amino-2,4-dibromoanthraquinone	+	81-49-2	-	*	*
Azoxymethane	+	25843-45-2	-	*	*
Chloroform	-	67-66-3	-	*	*
Cupferron	+	135-20-6	-	*	*
Dapsone	-	80-08-0	-	*	*
Dichloroacetylene	.	7572-29-4	-	*	*
Dimethyl methylphosphonate	-	756-79-6	-	*	*
Ethyl alcohol	-	64-17-5	-	*	*
N-Nitrosodiphenylamine	-	86-30-6	-	*	*
o-Phenylenediamine.2HCl	+	615-28-1	-	*	*
Z-ethyl-O,N,N-azoxyethane	.	16301-26-1	-	*	*
1,2-Dihydro-2-(5-nitro-2-thienyl)quinazolin-4(3H)-one	+	33389-33-2	-	1	0.99
1-Nitroso-1-hydroxyethyl-3-chloroethylurea	.	96806-34-7	-	0.981	0.981
2,4,5-Trimethylaniline	+	137-17-7	-	1	1
3-Amino-1,2,4-triazole	-	61-82-5	-	0.938	0.921
Acetaminophen	-	103-90-2	-	0.979	0.887

(table cont.)

Auramine O	+	2465-27-2	-	1	1
Dichlorvos	+	62-73-7	-	1	1
Hydantoin	ND	461-72-3	-	0.951	0.947
N-methyl-N'-nitro-N-nitrosoguanidine	+	70-25-7	-	0.917	0.917
N,N'-[6-(5-nitro-2-furyl)-s-triazine-2,4-diyl] bisacetamide	+	51325-35-0	-	0.963	0.959
Proflavine hydrochloride hemihydrate	+	952-23-8	-	1	1
Tris(2,3-dibromopropyl) phosphate	+	126-72-7	-	1	1
Trp-p-1 acetate	.	75104-43-7	-	0.99	0.992
2-Azoxypropane	.	----------	-	*	1
Chlorofluoromethane	-	593-70-4	-	*	*
Thioacetamide	-	62-55-5	-	*	*
Tris(2-chloroethyl)phosphate	-	115-96-8	-	*	*
Vinyl acetate	+	108-05-4	-	*	*
Vinyl bromide	.	593-60-2	-	*	*
1-(4-Chlorophenyl)-1-phenyl-2-propynyl carbamate	.	10473-70-8	-	*	*
2,6-Dinitrotoluene	+	606-20-2	-	*	*
5-Nitro-ortho-anisidine	-	99-59-2	-	*	*
DL-ethionine	-	67-21-0	-	*	*
Nitrobenzene	-	98-95-3	-	*	*
Trimethylthiourea	+	2489-77-2	-	*	*
N-methyl-N-nitrosobenzamide	+	63412-06-6	-	*	1
4-Chloro-o-phenylenediamine	-	95-83-0	-	*	*
Benzofuran	+	271-89-6	-	*	*
Nitroso-baygon	-	38777-13-8	-	*	*
2,3,7,8-Tetrachlorodibenzo-p-Dioxin	+	1746-01-6	-	*	*
2-Naphthylamine	.	91-59-8	-	*	*
beta-Butyrolactone	-	3068-88-0	-	*	*
Mirex photo	+	2385-85-5	-	*	*
p-Nitrosodiphenylamine	+	156-10-5	-	*	*
p-Quinone dioxime	+	105-11-3	-	*	*
Pararosaniline hydrochloride	+	569-61-9	-	*	*
Pentachloroanisole	-	1825-21-4	-	*	*
t-butyl alcohol	-	75-65-0	-	*	*
2,4,6-Trichlorophenol	+	88-06-2	-	*	0.091
Azaserine	.	115-02-6	-	*	0
Dehydroepiandrosterone	.	53-43-0	-	*	0
Tamoxifen citrate	+	54965-24-1	-	*	0
Urethane	.	51-79-6	-	*	0
Nitrosoamylurethane	+	64005-62-5	-	1	0.544
Nitrosoethylurethane	-	614-95-9	-	*	0.163
Propylthiouracil	+	51-52-5	-	*	0
Chrysazin	-	117-10-2	-	*	0.047
Diethylstilbestrol	+	56-53-1	-	*	0.045

(table cont.)

Furfural	-	98-01-1	-	0	0.9
Furan	+	110-00-9	-	0	0
p-Cresidine	-	120-71-8	-	0	0
Uracil	+	66-22-8	-	0	0
5-Azacytidine	.	320-67-2	-	0	0.109
4-(Methylnitrosamino)-1-(3-pyridyl)-1-butanol	+	--------	-	0.382	0.249
1,2-Epoxybutane	.	106-88-7	-	0	0.455
1-Nitroso-3,4,5-trimethylpiperazine	+	75881-18-4	-	0.024	0.032
1-Phenyl-3,3-dimethyltriazene	+	7227-91-0	-	0.02	0.03
4,4'-thiobisbenzenamine	.	139-65-1	-	0	0
6-Dimethylamino-4,4-diphenyl-3-heptanone hydrochloride	-	1095-90-5	-	0.028	0.037
Benzene	-	71-43-2	-	0	0
Chlorobenzene	+	108-90-7	-	0	0
C.I. Direct blue 6	+	2602-46-2	-	0.504	0.613
C.I. Direct black 38	.	1937-37-7	-	0.504	0.622
Ciprofibrate	ND	52214-84-3	-	0	0
Clofibride	+	26717-47-5	-	0	0
D&C red no. 5	.	3761-53-3	-	0	0.004
N,N'-dinitroso-perhydropyrimidine	-	15973-99-6	-	0.032	0.037
Ethinyl estradiol	-	57-63-6	-	0	0.219
FD&C red no.1	-	3564-09-8	-	0	0.004
FD&C red no. 2	.	915-67-3	-	0	0.005
Fumonisin B1	+	116355-83-0	-	0	0.051
Hydrazine sulfate	.	10034-93-2	-	0	0.038
ICRF 159	-	21416-87-5	-	0.059	0.061
Methimazole	-	60-56-0	-	0	0
Methyl t-butyl ether	+	1634-04-4	-	0	0
Mitomycin C	.	50-07-7	-	0.051	0.067
N-nitroso-bis-(4,4,4-trifluoro-n-butyl)amine	.	83335-32-4	-	0.022	0.032
N-Nitroso-N-methyldecylamine	+	75881-22-0	-	0.028	0.033
N-Nitrosodimethylamine	.	62-75-9	-	0.045	0.05
N-nitroso(methyl)-(2-hydroxyethyl)amine	+	26921-68-6	-	0.025	0.022
N-Nitrosopiperazine	.	5632-47-3	-	0.02	0.02
Nitroso-2-oxopropylethanolamine	+	92177-49-6	-	0.025	0.022
Nitrosoheptamethyleneimine	+	20917-49-1	-	0.022	0.032
o-Aminoazotoluene	-	97-56-3	-	0	0.023
Prednisolone	.	50-24-8	-	0	0
R-(-)-2-methyl-N-nitrosopiperidine	.	14026-03-0	-	0.02	0.027
Retinol acetate	.	127-47-9	-	0	0
Trenimon	.	68-76-8	-	0	0
Triamcinolone, acetonide	+	76-25-5	-	0	0
1,2-Dichloroethane	-	107-06-2	+	*	*

(table cont.)

Carbon tetrachloride	+	56-23-5	+	*	*
Methylene chloride	-	75-09-2	+	*	*
Nitromethane	+	75-52-5	+	*	*
2,2-Bis(bromomethyl)-1,3-propanediol, technical grade	+	3296-90-0	+	0.059	0.037
2,4-Dinitrotoluene (containing 1.0-1.5% 2,6-dinitrotoluene)	.	121-14-2	+	0	0
Acronycine	-	7008-42-6	+	0	0.357
Captafol	+	2425-06-1	+	0	0
Dibromomannitol	+	488-41-5	+	0.021	0.034
FD & C violet no. 1	-	1694-09-3	+	0.028	0.064
Ochratoxin A	-	303-47-9	+	0	0
Phenesterin	+	3546-10-9	+	0.013	0.025
Toluene diisocyanate, commercial grade (2,4 (80%)- and 2,6 (20%)	+	26471-62-5	+	0	0
2,4-Diaminoanisole sulfate	+	39156-41-7	+	*	*
Cytembena	+	16170-75-5	+	*	*
Glycidol	+	556-52-5	+	*	*
Propane sultone	+	1120-71-4	+	*	*
1,2-Propylene oxide	-	75-56-9	+	*	0
Acrylamide	+	79-06-1	+	*	0
1,3-Butadiene	-	106-99-0	+	*	*
1,4-Dioxane	+	123-91-1	+	*	*
Acrylonitrile	-	107-13-1	+	*	*
Chloroprene	+	126-99-8	+	*	*
o-Toluidine.HCl	+	636-21-5	+	*	*
Vinyl chloride	+	75-01-4	+	*	*
Sulfallate	+	95-06-7	+	0	0
Chlorambucil	+	305-03-3	+	1	0.343
Dacarbazine	+	4342-03-4	+	0.462	0.229
Phenacetin	+	62-44-2	+	*	0
1,2,3-Trichloropropane	+	96-18-4	+	*	*
1,2-Dibromo-3-chloropropane	+	96-12-8	+	*	*
1,2-Dibromoethane	+	106-93-4	+	*	*
2,4-Diaminotoluene	.	95-80-7	+	*	*
4,4'-Sulfonylbisacetanilide	.	77-46-3	+	*	*
Hexamethylmelamine	+	-------	+	*	*
4,4´-Methylene-bis(2-methylaniline)	+	838-88-0	+	0	0.792
Styrene	.	100-42-5	+	*	1
2-Methoxy-3-aminodibenzofuran	+	5834-17-3	+	*	1
4,4'-Methylene-bis(2-chloroaniline)	+	-------	+	*	1

(table cont.)

5-Nitro-2-furaldehyde semicarbazone	+	59-87-0	+	*	1
AF-2	-	3688-53-7	+	*	0.769
Procarbazine.HCl	+	366-70-1	+	*	1
trans-2-[(Dimethylamino)methylimino]-5-[2-(5-nitro-2-furyl)vinyl]-1,3,4-oxadiazole	+	55738-54-0	+	*	1
5-Nitroacenaphthene	+	602-87-9	+	0.846	0.923
1-(2-Hydroxyethyl)-1-nitrosourea	.	13743-07-2	+	0.914	0.93
1-[(5-Nitrofurfurylidene)amino]-2-imidazolidinone	+	555-84-0	+	0.942	0.935
1,2-Dimethyl-5-nitroimidazole	+	551-92-8	+	0.976	0.986
1-(2-Hydroxyethyl)-nitroso-3-ethylurea	.	-------	+	0.914	0.915
1,3-Dibutyl-1-nitrosourea	.	56654-52-5	+	0.932	0.926
1-Allyl-1-nitrosourea	.	760-56-5	+	0.909	0.963
1-Amyl-1-nitrosourea	.	10589-74-9	+	0.932	0.938
1-Ethylnitroso-3-(2-hydroxyethyl)-urea	.	-------	+	0.914	0.894
1-Ethylnitroso-3-(2-oxopropyl)-urea	.	-------	+	0.865	0.873
1-Nitropyrene	.	5522-43-0	+	1	1
2-(2,2-Dimethylhydrazino)-4-(5-nitro-2-furyl)thiazole	+	26049-69-4	+	0.974	0.97
2,2,2-Trifluoro-N-[4-(5-nitro-2-furyl)-2-thiazoly]acetamide	+	42011-48-3	+	0.965	0.961
2-Acetylaminofluorene	+	53-96-3	+	0.951	0.948
2-Amino-5-(5-nitro-2-furyl)-1,3,4-oxadiazole	.	3775-55-1	+	0.96	0.958
2-Amino-5-(5-nitro-2-furyl)-1,3,4-thiadiazole	.	712-68-5	+	0.93	0.929
2-Amino-5-nitrothiazole	+	121-66-4	+	0.982	0.988
2-Hydrazino-4-(5-nitro-2-furyl)thiazole	.	26049-68-3	+	0.974	0.971
2-Hydrazino-4-(*p*-aminophenyl)thiazole	.	26049-71-8	+	0.977	0.973
2-Hydrazino-4-(*p*-nitrophenyl)thiazole	.	26049-70-7	+	0.979	0.975
3-(5-Nitro-2-furyl)-imidazo(1,2-〈)pyridine	.	75198-31-1	+	0.938	0.932
3,3'-Dichlorobenzidine	+	91-94-1	+	0.941	0.923
3,3'-Dimethoxybenzidine.2HCl	+	20325-40-0	+	0.941	0.883
3,3'-Dimethylbenzidine.2HCl	+	612-82-8	+	0.941	0.923
3-Methylcholanthrene	+	56-49-5	+	0.895	0.722
4-(5-Nitro-2-furyl)thiazole	+	53757-28-1	+	0.97	0.965
4,6-Diamino-2-(5-nitro-2-furyl)-S-triazine	+	720-69-4	+	0.955	0.953
4,6-Dimethyl-2-(5-nitro-2-furyl)pyrimidine	.	59-35-8	+	0.918	0.92

(table cont.)

4-Acetylaminobiphenyl	.	4075-79-0	+	0.943	0.932
4-Aminodiphenyl.HCl	+	2113-61-3	+	0.941	0.931
4-Bis(2-hydroxyethyl)amino-2-(5-nitro-2-thienyl)quinazoline	-	33372-39-3	+	0.922	0.87
4-Methyl-1-[(5-nitrofurfurylidene)amino]-2-imidazolidinone	+	21638-36-8	+	0.919	0.873
Atrazine	-	1912-24-9	+	1	1
Bemitradine	.	88133-11-3	+	0.949	0.949
Benzidine	+	92-87-5	+	0.941	0.923
Carboxymethylnitrosourea	-	60391-92-6	+	0.909	0.963
Formic acid 2-(4-methyl-2-thiazolyl)hydrazide	.	32852-21-4	+	0.986	0.992
Formic acid 2-[4-(5-nitro-2-furyl)-2-thiazolyl]hydrazide	+	3570-75-0	+	0.968	0.965
Hydrazobenzene	+	122-66-7	+	1	1
Indomethacin	-	53-86-1	+	0.654	0.671
IQ	+	76180-96-6	+	0.979	0.987
IQ.HCl	+	-------	+	0.979	0.987
Isoniazid	+	54-85-3	+	0.913	0.937
l-5-Morpholinomethyl-3-[5(-nitrofurfurylidene)amino]-2-oxazolidinone.HCl	.	3031-51-4	+	0.797	0.634
Metronidazole	+	443-48-1	+	0.941	0.966
N-[4-(5-Nitro-2-furyl)-2-thiazolyl]acetamide	+	531-82-8	+	0.965	0.961
N-(9-Oxo-2-fluorenyl)acetamide	.	3096-50-2	+	0.943	0.932
N-(N-Methyl-N-nitrosocarbamoyl)-*l*-ornithine	.	63642-17-1	+	0.892	0.752
N-1-Diacetamidofluorene	.	63019-65-8	+	0.929	0.941
N-(2-Fluorenyl)-2,2,2-trifluoroacetamide	.	363-17-7	+	0.951	0.948
N-Hexylnitrosourea	+	18774-85-1	+	0.872	0.921
N,N'-[6-(5-Nitro-2-furyl)-s-triazine-2,4-diyl]bisacetamide	+	51325-35-0	+	0.917	0.918
N-*n*-Butyl-N-nitrosourea	+	869-01-2	+	0.932	0.938
Nithiazide	+	139-94-6	+	0.977	0.968
PhIP.HCl	+	--------	+	0.953	0.948
Trp-P-2-acetate	+	72254-58-1	+	0.913	0.927

Footnotes
. no *Salmonella* evaluation was made for the compound
* no prediction was made for the compound
ND: no *Salmonella* mutagenicity data available for compound
Lightly shaded background: correctly predicted noncarcinogen based on optimal cut-off value
Darkly shaded background: correctly predicted carcinogen based on optimal cut-off value
No shaded values: wrong prediction was made for compound

Table 4.6 Model validation for rat mammary gland carcinogens and non-carcinogens (MC-NC). For the ABC model, compounds with values above or equal to 37% were predicted to be active compounds and those below 37% were predicted to be inactive. For the ABCH model, compounds with values above or equal to 47% were predicted to be active and those below 47% were predicted as inactive.

Model 2

				Model 3-7/0.90	
Chemical	*Salmonella* Mutagenicity	CASN	Experimental Activity	ABC % Active (0.37)	ABCH % Active (0.47)
Dichlorodifluoromethane	.	75-71-8	-	*	*
Nitroethane	-	79-24-3	-	*	*
Tetrafluoro-m-phenylenediamine dihydrochloride	.	63886-77-1	-	*	*
1-Phenyl-2-thiourea	-	103-85-5	-	*	*
1-Chloro-2-nitrobenzene	+	88-73-3	-	*	*
Acetaldoxime	-	107-29-9	-	*	1
Benzaldehyde	-	100-52-7	-	*	0.9
m-Toluidine hydrochloride	-	638-03-9	-	*	0.923
HC blue no. 2	+	33229-34-4	-	*	1
1-Nitroso-5,6-dihydrothymine	+	62641-67-2	-	0.882	0.889
3-Hydroxy-4-acetylaminobiphenyl	.	4463-22-3	-	0.97	0.925
5-Nitro-2-furamidoxime	+	772-43-0	-	0.947	0.94
5-Nitro-2-furanmethanediol diacetate	+	92-55-7	-	0.743	0.715
Caffeine	-	58-08-2	-	0.789	0.837
Diazepam	-	439-14-5	-	0.909	0.909
Formic acid 2-[4-(2-furyl)-2-thiazolyl]hydrazide	.	31873-81-1	-	0.949	0.943
Norharman	+	244-63-3	-	1	1
Pyrimethamine	-	58-14-0	-	0.959	0.911
1-(2-Hydroxypropyl)-nitroso-3-chloroethylurea	.	96806-35-8	-	0.948	0.94
Benzylthiocyanate	.	3012-37-1	-	*	*
Methyl Parathion	+	298-00-0	-	*	*
1,1,2,2-Tetrachloroethane	-	79-34-5	-	*	*
N-Methylolacrylamide	-	924-42-5	-	*	1
Dioxathion	+	78-34-2	-	*	*
Dithiooxamide	.	79-40-3	-	*	*
Pentaerythritol tetranitrate	-	78-11-5	-	*	*
p,p'-DDE	-	72-55-9	-	*	*
S-ethyl-L-cysteine	.	2629-59-6	-	*	1
Tace	-	569-57-3	-	*	*
trans-Anethole	.	4180-23-8	-	*	0.815
3-Nitrosomethylaminopyridine	-	69658-91-9	-	*	0
Propylene	+	115-07-1	-	*	0
Malathion	-	121-75-5	-	*	0
Methyl carbazate	.	6294-89-9	-	*	0
O,O-diethyl O-(3,5,6-trichloro-2-pyridinyl) phosphorothioate	-	2921-88-2	-	*	0
Phenethyl isothiocyanate	.	2257-09-2	-	*	0.1
Clonitralid	.	1420-04-8	-	*	0.077

(table cont.)

Sulfisoxazole	-	127-69-5	-	0	0.816
C.I. Pigment yellow 12	-	6358-85-6	-	0	0.63
Ipazilide fumarate	.	115436-74-3	-	0.083	0.62
C.I. Orange 10	-	1936-15-8	-	0	0.029
FD&C Yellow No. 6	-	2783-94-0	-	0	0.04
Acrylic acid	-	79-10-7	-	0	0
1-O-Hexyl-2,3,5-trimethylhydroquinone (HTHQ)	.	148081-72-5	-	0.035	0.036
2,2'-Methylenebis(4-methyl-6-tert-butylphenol)	.	119-47-1	-	0	0.089
2,7-Dichlorodibenzo-p-dioxin (DCDD)	-	33857-26-0	-	0	0.03
2-Chloroacetophenone	-	532-27-4	-	0	0
2-Difluoromethylornithine	.	70052-12-9	-	0	0
4-Nitroanthranilic acid	+	619-17-0	-	0.098	0.111
5-Fluorouracil	-	51-21-8	-	0	0
6-Aminocaproic acid	.	60-32-2	-	0	0.2
alpha-Methyldopa sesquihydrate	-	41372-08-1	-	0.038	0.034
Benzyl alcohol	-	100-51-6	-	0	0
Butylated hydroxytoluene	-	128-37-0	-	0	0.031
Butyl benzyl phthalate	-	85-68-7	-	0.039	0.017
Carbromal	-	77-65-6	-	0.36	0.468
Codeine	-	76-57-3	-	0.02	0.027
Compound 50-892	.	65765-07-3	-	0	0.056
Cyclohexanone	-	108-94-1	-	0	0
Dicofol	-	115-32-2	-	0	0
Dicyclopentadiene dioxide	.	81-21-0	-	0.032	0.032
Diphenhydramine hydrochloride	-	147-24-0	-	0	0.02
Dipyrone	.	68-89-3	-	0.143	0.3
DL-alpha-tocopheryl acetate	-	58-95-7	-	0.031	0.039
DL-Menthol	-	15356-70-4	-	0.044	0.051
Dopamine HCl	+	62-31-7	-	0.044	0.038
DL-diepoxybutane	+	298-18-0	-	0	0
Ellagic acid	-	476-66-4	-	0.026	0.03
Endosulfan	-	115-29-7	-	0.038	0.044
Endrin	-	72-20-8	-	0.03	0.032
Ephedrine sulfate	-	134-72-5	-	0	0
Erythromycin stearate	-	643-22-1	-	0.022	0.028
FD&C Red No. 3	-	16423-68-0	-	0.042	0.046
Fenaminosulf (formulated)	+	140-56-7	-	0	0
Fenvalerate	.	51630-58-1	-	0.046	0.028
Gemfibrozil	.	25812-30-0	-	0.035	0.038
Heptylamine	.	111-68-2	-	0.07	0.063
Hexachlorocyclopentadiene (HCCPD)	-	77-47-4	-	0	0
Iodoacetamide	.	144-48-9	-	0	0
Kaempferol	+	520-18-3	-	0.027	0.033
Monochloroacetic acid	-	79-11-8	-	0	0
N-methyldopamine,O,O'-diisobutyroyl ester hydrochloride	.	75011-65-3	-	0.031	0.025
N-Methyl-2-Pyrrolidone	-	872-50-4	-	0.04	0.055
N,N-dipropyl-4-(4'-[pyridyl-1'-oxide]azo)aniline	.	------	-	0.022	0.054

(table cont.)

Orotic acid, monosodium salt	.	154-85-8	-	0	0
Oxprenolol hydrochloride	.	6452-73-9	-	0.01	0.031
Oxytetracycline hydrochloride	-	2058-46-0	-	0.023	0.03
O,S-dibenzoyl thiamine hydrochloride	.	35660-60-7	-	0.14	0.424
Picloram (technical grade)	-	1918-02-1]	-	0.071	0.059
Pimaricin	.	7681-93-8	-	0.012	0.017
Piperonyl sulfoxide	-	120-62-7	-	0.044	0.036
Probenecid	-	57-66-9	-	0.091	0.07
Propyl gallate	-	121-79-9	-	0.036	0.032
Rutin sulfate	.	12768-44-4	-	0.019	0.023
Sodium bicarbonate	.	144-55-8	-	0	0
Sotalol hydrochloride	.	959-24-0	-	0.111	0.06
Tetracycline.HCl	-	64-75-5	-	0.024	0.032
Triprolidine.HCl monohydrate	-	6138-79-0	-	0.056	0.08
Zatosetron maleate	.	123482-22-4	-	0.065	0.076
Zearalenone	-	17924-92-4	-	0.055	0.041
1,2-Propylene oxide		75-56-9	+	*	0
2,2-Bis(bromomethyl)-1,3-propanediol, technical grade		3296-90-0	+	0.034	0.175
2,4-Diaminoanisole sulfate		39156-41-7	+	*	*
4,4'-Methylene-bis(2-chloroaniline)		101-14-4	+	*	*
Acronycine		7008-42-6	+	0	0
Acrylamide		79-06-1	+	*	*
Acrylonitrile		107-13-1	+	*	*
Captafol		2939-80-2	+	0.021	0.027
Carbon tetrachloride		56-23-5	+	*	*
Chlorambucil		305-03-3	+	0.074	0.062
Cytembena		16170-75-5	+	0	0
Dibromomannitol		488-41-5	+	0.032	0.096
FD & C violet no. 1		1694-09-3	+	0	0
Glycidol		556-52-5	+	*	0
Isoniazid		54-85-3	+	0	0
Methylene chloride		75-09-2	+	*	*
Nitromethane		75-52-5	+	*	*
Ochratoxin A		303-47-9	+	0.043	0.032
Phenesterin		3546-10-9	+	0.012	0.022
Procarbazine.HCl		366-70-1	+	*	*
Propane sultone		1120-71-4	+	*	*
Styrene		100-42-5	+	*	0
Sulfallate		95-06-7	+	0	0
Toluene diisocyanate, commercial grade (2,4 (80%)- and 2,6 (20%)		584-84-9	+	0	0
Vinyl chloride		75-01-4	+	*	*
1,4-Dioxane		123-91-1	+	0	0
1,2-Dichloroethane		107-06-2	+	*	*
1,3-Butadiene		106-99-0	+	*	*
Chloroprene		126-99-8	+	*	*
Hydrazobenzene		122-66-7	+	*	*
N-(N-Methyl-N-nitrosocarbamoyl)-*l*-ornithine		63642-17-1	+	0.1	0.068
Dacarbazine		4342-03-4	+	0	0

(table cont.)

2-Methoxy-3-aminodibenzofuran	5834-17-3	+	*	*
Metronidazole	443-48-1	+	*	*
1,2-Dimethyl-5-nitroimidazole	551-92-8	+	*	*
2,4-Dinitrotoluene (containing 1.0-1.5% 2,6-dinitrotoluene)	121-14-2	+	0	0
Phenacetin	62-44-2	+	*	0.633
1-Ethylnitroso-3-(2-hydroxyethyl)-urea	-------	+	*	0.557
4,4'-Sulfonylbisacetanilide	77-46-3	+	*	1
1,2,3-Trichloropropane	96-18-4	+	*	1
1,2-Dibromoethane	106-93-4	+	*	1
1,2-Dibromo-3-chloropropane	96-12-8	+	*	1
1-(2-Hydroxyethyl)-nitroso-3-ethylurea	-------	+	*	0.98
1-(2-Hydroxyethyl)-1-nitrosourea	13743-07-2	+	*	0.989
3,3'-Dimethoxybenzidine.2HCl	20325-40-0	+	0.9	0.414
Indomethacin	53-86-1	+	1	0.806
l-5-Morpholinomethyl-3-[(5-nitrofurfurylidene)amino]-2-oxazolidinone.HCl	3031-51-4	+	0.736	0.544
3,3'-Dichlorobenzidine	91-94-1	+	0.9	0.913
N-(9-Oxo-2-fluorenyl)acetamide	3096-50-2	+	0.9	0.957
4-Bis(2-hydroxyethyl)amino-2-(5-nitro-2-thienyl)quinazoline	33372-39-3	+	0.94	0.776
Atrazine	1912-24-9	+	1	1
N-Hexylnitrosourea	18774-85-1	+	0.722	0.95
1-(2-Hydroethyl)-3-[(5-nitrofurfurylidene)amino]-2-imidazolidinone	-------	+	1	0.959
1,3-Dibutyl-1-nitrosourea	56654-52-5	+	1	0.987
1-Allyl-1-nitrosourea	760-56-5	+	0.667	0.891
1-Amyl-1-nitrosourea	10589-74-9	+	1	0.991
1-Ethylnitroso-3-(2-oxopropyl)-urea	110559-84-7	+	1	0.983
1-Nitropyrene	5522-43-0	+	1	1
1-[(5-Nitrofurfurylidene)amino]-2-imidazolidinone	555-84-0	+	1	0.999
2,2,2-Trifluoro-N-[4-(5-nitro-2-furyl)-2 thiazolyl]acetamide	42011-48-3	+	0.935	0.921
2,4-Diaminotoluene	95-80-7	+	1	1
2-Acetylaminofluorene	53-96-3	+	0.9	0.957
2-Amino-5-nitrothiazole	121-66-4	+	0.97	0.967
2-Amino-5-(5-nitro-2-furyl)-1,3,4-oxadiazole	3775-55-1	+	1	1
2-Amino-5-(5-nitro-2-furyl)-1,3,4-thiadiazole	712-68-5	+	0.992	0.994
2-Hydrazino-4-(5-nitro-2-furyl)thiazole	26049-68-3	+	0.994	0.995
2-Hydrazino-4-(*p*-aminophenyl)thiazole	26049-71-8	+	0.986	0.988
2-Hydrazino-4-(*p*-nitrophenyl)thiazole	26049-70-7	+	0.99	0.992
2-(2,2-Dimethylhydrazino)-4-(5-nitro-2-furyl)thiazole	26049-69-4	+	0.994	0.995
3,3'-Dimethylbenzidine.2HCl	612-82-8	+	0.944	0.964
3-Methylcholanthrene	56-49-5	+	0.915	0.858
3-(5-Nitro-2-furyl)imidazo(1,2-α)pyridine	------	+	1	1

(table cont.)

4,4'-Methylene-bis(2-methylaniline)	838-88-0	+	0.667	0.784
4,6-Diamino-2-(5-nitro-2-furyl)-S-triazine	720-69-4	+	1	1
4,6-Dimethyl-2-(5-nitro-2-furyl)pyrimidine	59-35-8	+	1	1
4-Acetylaminobiphenyl	4075-79-0	+	0.9	0.957
4-Aminodiphenyl.HCl	2113-61-3	+	0.9	0.913
4-Methyl-1-[(5-nitrofurfurylidene)amino]-2-imidazolidinone	21638-36-8	+	1	0.996
4-(5-Nitro-2-furyl)thiazole	53757-28-1	+	0.997	0.998
5-Nitroacenaphthene	602-87-9	+	1	1
5-Nitro-2-furaldehyde semicarbazone	59-87-0	+	1	1
AF-2	3688-53-7	+	1	1
Bemitradine	88133-11-3	+	0.996	0.996
Benzidine	92-87-5	+	0.9	0.913
Carboxymethylnitrosourea	60391-92-6	+	1	1
Formic acid 2-[4-(5-nitro-2-furyl)-2-thiazolyl]hydrazide	3570-75-0	+	0.996	0.996
Formic acid 2-(4-methyl-2-thiazolyl)hydrazide	32852-21-4	+	0.975	0.98
Hexamethylmelamine	645-05-6	+	1	1
IQ	76180-96-6	+	1	1
IQ.HCl	76180-96-6	+	1	1
Nithiazide	139-94-6	+	0.97	0.967
N-1-Diacetamidofluorene	63019-65-8	+	1	0.947
N-(2-Fluorenyl)-2,2,2-trifluoroacetamide	363-17-7	+	0.9	0.945
N,N'-[6-(5-Nitro-2-furyl)-s-triazine-2,4-diyl]bisacetamide	51325-35-0	+	1	1
N-*n*-Butyl-N-nitrosourea	869-01-2	+	1	0.991
N-[4-(5-Nitro-2-furyl)-2-thiazolyl]acetamide	531-82-8	+	0.996	0.996
o-Toluidine.HCl	636-21-5	+	1	1
PhIP.HCl	105650-23-5	+	1	1
trans-2-[(Dimethylamino)methylimino]-5-[2-(5-nitro-2-furyl)vinyl]-1,3,4-oxadiazole	55738-54-0	+	1	1
Trp-P-2-acetate	72254-58-1	+	0.778	0.914

Footnotes: See Table 4.5

Table 4.7 Model validation for mouse mammary gland carcinogens and non-mammary gland carcinogens (MC-NMC). For the ABC 3/0.75 model, compounds with values equal to or above 76% were predicted to be active compounds and those below 76% were predicted to be inactive. For the ABCH 3/0.90, ABC 3/0.90 and ABCH 3/0.75 models, the optimal cut-off values were 62%, 95% and 76%, respectively.

Model 2			ABC 3/0.90	ABCH 3/0.90	ABC 3/0.75	ABCH 3/0.75
Chemical	CASN	Experimental Activity	% Active (0.95)	% Active (0.62)	% Active (0.76)	% Active (0.76)
Cyclamate sodium	100-88-9	-	*	1	0.808	0.853
Daminozide	1596-84-5	-	1	1	1	0.934
Dichloroacetylene	7572-29-4	-	*	*	*	*
Isoprene	78-79-5	-	1	0.985	0.95	0.921
Trichloroaceticacid	76-03-9	-	*	*	*	*
Tris(2,3-dibromopropyl) phosphate	126-72-7	-	0.944	0.98	0.944	0.896
Ethylhydrazine.HCl	18413-14-4	-	*	*	0.75	0.762
1,1,2,2-Tetrachloroethane	79-34-5	-	*	*	0.25	0.227
1,2-Dichloropropane	78-87-5	-	*	1	0.25	0.67
alpha-1,2,3,4,5,6-Hexachlorocyclohexane-d6	86194-41-4	-	*	*	0.731	0.757
1'-Hydroxysafrole	5208-87-7	-	*	0.618	0.228	0.33
3-Methoxy-4-aminoazobenzene	3544-23-8	-	*	0.265	0.182	0.207
3-(Chloromethyl) pyridine hydrochloride	6959-48-4	-	*	0.333	0.244	0.209
4-Chloro-4'-aminodiphenylether	101-79-1	-	*	0.1	0.182	0.176
4-Chloro-*m*-phenylenediamine	5131-60-2	-	*	0.1	0.182	0.169
Benzyl chloride	100-44-7	-	*	0	0.178	0.181
m-Toluidine.HCl	638-03-9	-	*	0.077	0.18	0.173
p,p'-DDD	72-54-8	-	*	0	0.182	0.183
2-Acetylaminofluorene	53-96-3	-	0	0.059	0.183	0.183
2-Hydrazino-4-(p-aminophenyl)thiazole	26049-71-8	-	0	0.059	0.187	0.177
3,3',4,4'-Tetraaminobiphenyl tetrahydrochloride	7411-49-6	-	0	0.059	0.185	0.179
3,3'-Dimethylbenzidine dihydrochloride	612-82-8	-	0	0.059	0.183	0.181
Benzidine.2HCl	531-85-1	-	0	0.059	0.185	0.178
Phenacetin	62-44-2	-	0.9	0.486	0.232	0.274
1,2-Dibromoethane	106-93-4	+	*	*	*	*
1,2-Dichloroethane	107-06-2	+	0	0	0	0
Benzene	71-43-2	+	*	*	*	*
C.I. Direct black 38	1937-37-7	+	0.046	0.057	0.046	0.059
Furosemide	54-31-9	+	0.426	0.289	0.372	0.257
Nitrobenzene	98-95-3	+	*	*	*	*
Estradiol	50-28-2	+	0.551	0.681	0.524	0.608

(table cont.)

Vinyl fluoride	75-02-5	+	*	*	*	0.769
1,3-Butadiene	106-99-0	+	*	*	0.8	0.78
5-Azacytidine	320-67-2	+	1	1	0.955	0.943
α-Ecdysone	3604-87-3	+	1	1	0.96	0.949
Calciferol	50-14-6	+	1	1	0.948	0.931
Chloroprene	126-99-8	+	1	1	0.875	0.82
Diethylstilbestrol	56-53-1	+	1	1	0.986	0.976
Ethylene oxide	75-21-8	+	1	1	0.857	0.844
Glycidol	556-52-5	+	1	1	0.929	0.939
Griseofulvin	126-07-8	+	1	0.987	0.77	0.808
Isoniazid	54-85-3	+	1	0.975	1	0.888
Isonicotinic acid vanillylidenehydrazide	149-17-7	+	1	1	1	0.928
(N-6)-(Methylnitroso)adenosine	------	+	1	1	0.947	0.964
Reserpine	50-55-5	+	1	1	0.95	0.926
Sulfallate	95-06-7	+	1	1	1	0.829
Vinylidene chloride	75-35-4	+	1	1	1	0.829
Vinyl chloride	75-01-4	+	1	1	1	0.829

Footnotes: See Table 4.5

Table 4.8 Model validation for mouse mammary gland carcinogens and rodent non-carcinogens (MC-RNC). For the ABC 3/0.75 model, compounds with values equal to or greater than 30% were predicted to be active compounds and those below 30% were predicted to be inactive. For the ABCH 3/0.90, ABC 3/0.90, and ABCH 3/0.75 models, these optimal cut-off values were 93%, 1% and 42%, respectively.

Model 1			ABC 3/0.90	ABCH 3/0.90	ABC 3/0.75	ABCH 3/0.75
Chemical	CASN	Experimental Activity	% Active (0.01)	% Active (0.93)	% Active (0.3)	% Active (0.42)
1,1,1-Trichloroethane (technical grade)	71-55-6	-	*	*	*	*
3-Sulfolene	77-79-2	-	*	*	*	0.8
Benzoguanamine	91-76-9	-	*	*	*	*
Cyanamide, calcium	156-62-7	-	*	*	*	*
Dichlorodifluoromethane	75-71-8	-	*	*	*	*
Hexamethylenetetramine	100-97-0	-	*	*	*	*
Octachlorostyrene	29082-74-4	-	1	1	1	1
Propylene	115-07-1	-	*	1	*	0.875
Saccharin	81-07-2	-	1	1	0.78	0.857
Deltamethrin	52918-63-5	-	1	0.919	0.459	0.556
Gemfibrozil	25812-30-0	-	1	0.925	0.534	0.659
Hydrochlorothiazide	58-93-5	-	*	*	0.25	0.25
Pyrazinamide	98-96-4	-	*	*	0.25	0.412
Urea	57-13-6	-	*	*	0.25	0.25

(table cont.)

6-Dimethylamino-4,4-diphenyl-3-heptanone hydrochloride	1095-90-5	-	*	0	0.176	0.291
Dimethylformamide	68-12-2	-	*	0	0.222	0.167
Methoxychlor	72-43-5	-	*	0	0.298	0.294
Omeprazole	73590-58-6	-	1	0.167	0.234	0.284
Oxamyl	23135-22-0	-	*	0	0.222	0.167
Tolbutamide	64-77-7	-	*	0	0.188	0.21
Caffeine	58-08-2	-	0	0	0.163	0.156
Diazepam	439-14-5	-	0	0	0.163	0.221
Nefiracetam	77191-36-7	-	0	0.108	0.168	0.274
Prazepam	2955-38-6	-	0	0.526	0.163	0.414
1,2-Dibromoethane	106-93-4	+	*	*	*	*
1,2-Dichloroethane	107-06-2	+	*	*	*	*
Benzene	71-43-2	+	*	*	*	*
Nitrobenzene	98-95-3	+	*	*	*	*
Furosemide	54-31-9	+	0	0	0.165	0.146
Sulfallate	95-06-7	+	0	0.176	0.24	0.268
Vinyl fluoride	75-02-5	+	*	*	*	0.833
Isoniazid	54-85-3	+	*	*	0.429	0.465
C.I. Direct black 38	1937-37-7	+	*	1	0.221	0.265
Vinylidene chloride	75-35-4	+	*	1	0.75	0.875
Vinyl chloride	75-01-4	+	*	1	0.75	0.857
Estradiol	50-28-2	+	0.442	0.617	0.484	0.6
Griseofulvin	126-07-8	+	0.432	0.569	0.447	0.577
5-Azacytidine	320-67-2	+	0.719	0.869	0.615	0.794
1,3-Butadiene	106-99-0	+	1	1	1	0.9
α-Ecdysone	3604-87-3	+	1	0.988	0.865	0.893
Calciferol	50-14-6	+	1	0.966	0.924	0.888
Chloroprene	126-99-8	+	1	1	0.875	0.9
Diethylstilbestrol	56-53-1	+	1	1	0.923	0.848
Ethylene oxide	75-21-8	+	1	1	0.923	0.9
Glycidol	556-52-5	+	1	1	0.951	0.954
Isonicotinic acid vanillylidenehydrazide	149-17-7	+	1	1	0.508	0.522
(N-6)-(Methylnitroso) adenosine	------	+	1	0.987	0.887	0.913
Reserpine	50-55-5	+	1	0.964	0.787	0.781

Footnotes: See Table 4.5

4.1.3 Predictive Performance of cat-SAR CPDB Rodent Models

The models assessed for predictivity were the general rat and mouse, and female-specific rodent ABC and ABCH models. During LOO-CV the sensitivity, specificity, and concordance values of the models were established. Through LOO-CV, both rodent and female-specific models produced results in favor of the ABC 90% requirement. The rat

model attained a sensitivity of 70% and a specificity of 69% yielding a concordance of 70% (ABC 3/0.90, Table 4.9). Predictions were made on 459 of the chemicals in the learning set. This learning set was comprised of a total of 946 chemicals. Therefore, the model made predictions on about half (i.e., 49%) of the test compounds. Likewise, the general mouse model was able to achieve a concordance between experimental and predicted results of 70% with a sensitivity of 74% and a specificity of 67% (ABC 3/0.90, Table 4.10). For this model, predictions were made on 346 of the 769 chemicals in the learning set. Again, the model was able to predict about half (i.e., 45%) of the test compounds in the dataset.

Table 4.9 Predictive performance of the general CPDB rat carcinogen model with 3 to 7 heavy atoms.

Model (opt. 0.27)	Total Fragments	Model Fragments	Active Fragments	Inactive Fragments	Sensitivity	Specificity	OCP
ABC 3/0.75	41886	5026	2450	2576	0.61(235/386)	0.74(257/349)	0.67(492/735)
ABC 3/0.90	**41886**	**3400**	**1773**	**1627**	**0.70(171/243)**	**0.69(148/216)**	**0.70(319/459)**
ABCH 3/0.75	89509	11716	5496	6220	0.62(267/434)	0.73(290/398)	0.67(557/832)
ABCH 3/0.90	89509	7594	3885	3709	0.63(199/317)	0.73(213/292)	0.68(412/609)

Footnotes: See Table 4.1

Table 4.10 Predictive performance of the general CPDB mouse carcinogen model with 3 to 7 heavy atoms.

Model (opt. 0.51)	Total Fragments	Model Fragments	Active Fragments	Inactive Fragments	Sensitivity	Specificity	OCP
ABC 3/0.75	33560	4070	1282	2788	0.60(177/293)	0.72(207/289)	0.66(384/582)
ABC 3/0.90	**33560**	**2428**	**919**	**1509**	**0.74(121/163)**	**0.67(122/183)**	**0.70(243/346)**
ABCH 3/0.75	70076	10070	3268	6802	0.69(237/342)	0.60(207/348)	0.64(444/690)
ABCH 3/0.90	70076	6027	2369	3658	0.66(159/242)	0.73(175/241)	0.69(334/483)

Footnotes: See Table 4.1

Analysis of the female-specific rodent datasets also provided findings reflective of the model's consistency. Taking into consideration that these two models were subsets of the general rodent models, it was expected that the predictive performance of these models would be similar to that of the general rodent models. The best female rat model achieved a sensitivity of 61% and a specificity of 73% yielding an OCP of 67% (ABC 3/0.90, Table 4.11). This learning set was comprised of 723 chemicals and the model made predictions on 357 test compounds. Therefore, the model made predictions on approximately 49% of the compounds in the learning set. Lastly, the best female mouse model had an OCP of 73% with a sensitivity of 61%, and a specificity of 83% (ABC 3/0.90, Table 4.12). This model made predictions on 371 of its 738 test compounds (i.e., 50% of all chemicals comprising the model).

Table 4.11 Predictive performance of the CPDB female rat carcinogen model with 3 to 7 heavy atoms and 75% and 90% optimal cut-offs of 0.26 and 0.32, respectively.

Model (opt. 0.32)	Total Fragments	Model Fragments	Active Fragments	Inactive Fragments	Sensitivity	Specificity	OCP
ABC 3/0.75	36131	3793	1500	2293	0.69(203/293)	0.64(171/269)	0.67(374/562)
ABC 3/0.90	**36131**	**2557**	**1062**	**1495**	**0.61(108/176)**	**0.73(132/181)**	**0.67(240/357)**

Footnotes: See Table 4.1

Table 4.12 Predictive performance of the CPDB female mouse carcinogen model with 3 to 7 heavy atoms and 75% and 90% optimal cut-offs of 0.48 and 0.51, respectively.

Model (opt. 0.51)	Total Fragments	Model Fragments	Active Fragments	Inactive Fragments	Sensitivity	Specificity	OCP
ABC 3/0.75	32925	3339	1076	2263	0.60(140/234)	0.72(234/290)	0.71(374/524)
ABC 3/0.90	**32925**	**2617**	**804**	**1813**	**0.61(102/167)**	**0.83(169/204)**	**0.73(271/371)**

Footnotes: See Table 4.1

4.1.4 Analysis of the General Rodent Models

Of importance, all four mammary and female-specific learning sets were subsets of the CPDB general rat and mouse learning sets. Model consistency in terms of OCP values was observed among ABC and ABCH models. It was also noted that the 75% models made predictions on a greater number of test compounds as previously seen with the mammary carcinogen models. For example, when comparing the 75% and 90% proportions for the rat dataset, the ABC 75% model made predictions on 735 chemicals in the dataset, whereas the ABC 90% model made predictions on 459 test chemicals in the dataset (Table 4.10). In other words, the rat ABC 75% model was able to make predictions on 78% of the 946 chemicals in the learning set compared to 49% by the ABC 90% model.

The best model produced for the mouse-female dataset was from the ABC 90% criterion. It is important to note that although this model produced a high predictivity, it was not the more inclusive model in comparison to the ABC 75% (Table 4.9). However, for the rat female dataset, the most predictive model was that of the ABC 75% criterion. This model was chosen for its predictive capacity to correctly predict a great number of active and inactive chemicals. Again, when comparing the ABC and ABCH models it was observed that the ABCH increased ability to make predictions on a large number of chemicals did not improve the model's OCP value. Instead, the model's accuracy remained almost stable and in some cases the OCP was slightly reduced.

To judge the predictive performance of the cat-SAR system, the NTP *Salmonella* mutagenicity database, and an analysis of the CPDB by Gold and collaborators were considered as 'gold-standards'. The rodent results were also compared to that of a separate study published on organ-specific carcinogenic databases by Young and others (Young et al

2004). Young *et al.* SAR analyses based on chemicophysical parameters for validating rodent carcinogens, the model was at best 30% sensitive and 77% specific (Young et al 2004). This uneven distribution of correct active and inactive predictions appears to be a common trend seen in current SAR approaches. However, balanced sensitivity and specificity percent values as presented in the previous tables (Table 4.9-4.12) are characteristic of the cat-SAR models.

Another published CPDB rat and mouse SAR study using the CASE/MULTICASE (MCASE) expert system by Cunningham *et al.* achieved concordances similar to that of the cat-SAR rat and mouse models. In a 10-fold cross-validation study where 10 disjoint sets of 10% of the chemicals were removed from the rat database and the remaining 90% of the chemicals were used as a learning set, MCASE was able to achieve a concordance between experimental and predicted results of 64% (Cunningham et al 1998). The model's sensitivity and specificity values were 55% and 73%, respectively. However, when Cunningham *et al.* applied a modified validation process (i.e., removal of all chemicals that were identified by a 'unique' structural alert were removed from consideration) designed to investigate the predictivity of a more focused rat SAR model, MCASE achieved a concordance of 71% (Cunningham et al 1998). In light of the model's respectable OCP value, the model's sensitivity and specificity values attained were 69% and 73%, respectively. In a similar MCASE study designed to identify genotoxic and non-genotoxic alerts for cancer in mice using the CPDB data, a concordance of 70% with a sensitivity of 63% and a specificity of 78% was achieved through LOO-CV. These values are close to the overall rat and mouse concordances derived from the cat-SAR study and indicate that the cat-SAR and MCASE systems are performing similarly.

Furthermore, Gold *et al.* stated that in 'near replicate' (i.e., cancer bioassay repeated with the same chemical) comparisons of rat cancer bioassays a reproducibility of 85% is estimated while mouse bioassays are estimated to be 80% reproducible (Gold 1987). The *Salmonella* mammalian microsome mutagenicity (Ames) test was designed to measure mutations using several strains of the *Salmonella typhimurium* (Ames et al 1973). As stated, the interlaboratory reproducibility of the *Salmonella* assay is 85% (Zeiger 1985). The cat-SAR system achieved an estimated 80% concordance for the *Salmonella* mutagenicity assay (data not shown). Taking into consideration the complexity of the endpoint (i.e., mechanisms of carcinogenesis), the cat-SAR system can usefully complement the results of the rodent bioassay and short-term tests.

4.1.5 Training/Test Set: Respiratory Sensitization cat-SAR Study

The cat-SAR method has been employed successfully in validating and predicting the biological activity of human respiratory chemical sensitizers (Cunningham et al; in press 2005). This group of compounds served as a test set in developing the cat-SAR method and have been accepted for publication. Once again, in comparison to results attained from other SAR approaches, the respiratory sensitization cat-SAR models were very successful in terms of its sensitivity, specificity, and OCP (Table 4.13). These analyses also illustrate that cat-SAR is applicable to diverse biological endpoints and that the predictive performance of the rodent and mammary-specific carcinogenicity models was not attained by chance. It should be noted that it is speculated that the respiratory sensitization ABC and ABCH models demonstrated a better predictivity because they were limited in the number of associated mechanisms.

Table 4.13 Predictive performance of ABC and ABCH respiratory sensitization models based on fragments of size between three and seven heavy atoms and considered atoms, bonds, and atom connection. The ABCH model also included consideration of hydrogen atoms.

Model	Total Fragments	Model Fragments	Active Fragments	Inactive Fragments	Sensitivity	Specificity	OCP
ABC	5737	1305	1213	92	0.94	0.87	0.91
ABCH	14424	3356	2926	430	0.89	0.95	0.92

Footnotes: See Table 4.1

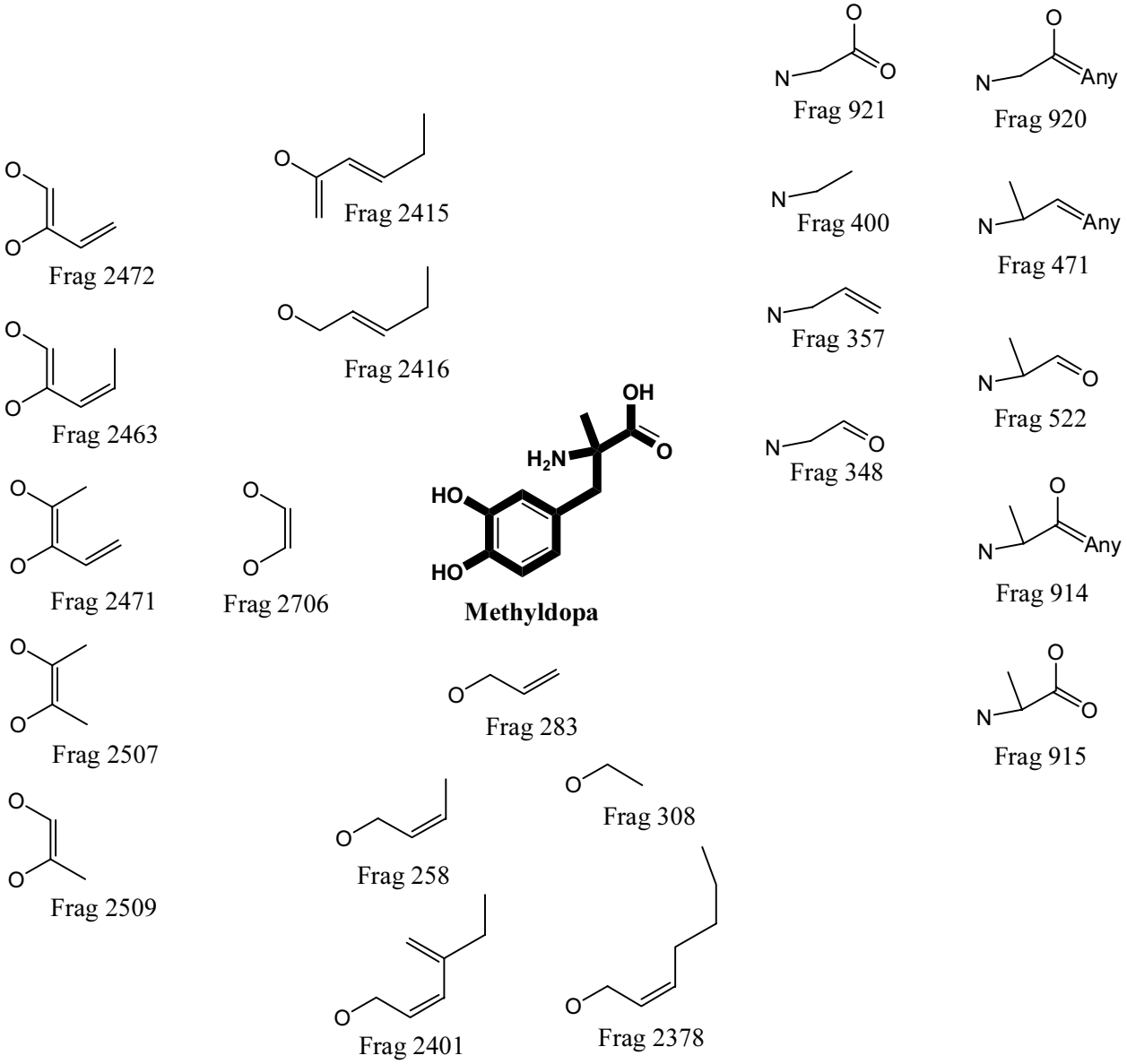

Figure 4.1 Illustration of the 22 significant fragments contributing to the active validation prediction of the respiratory sensitizer methyldopa.

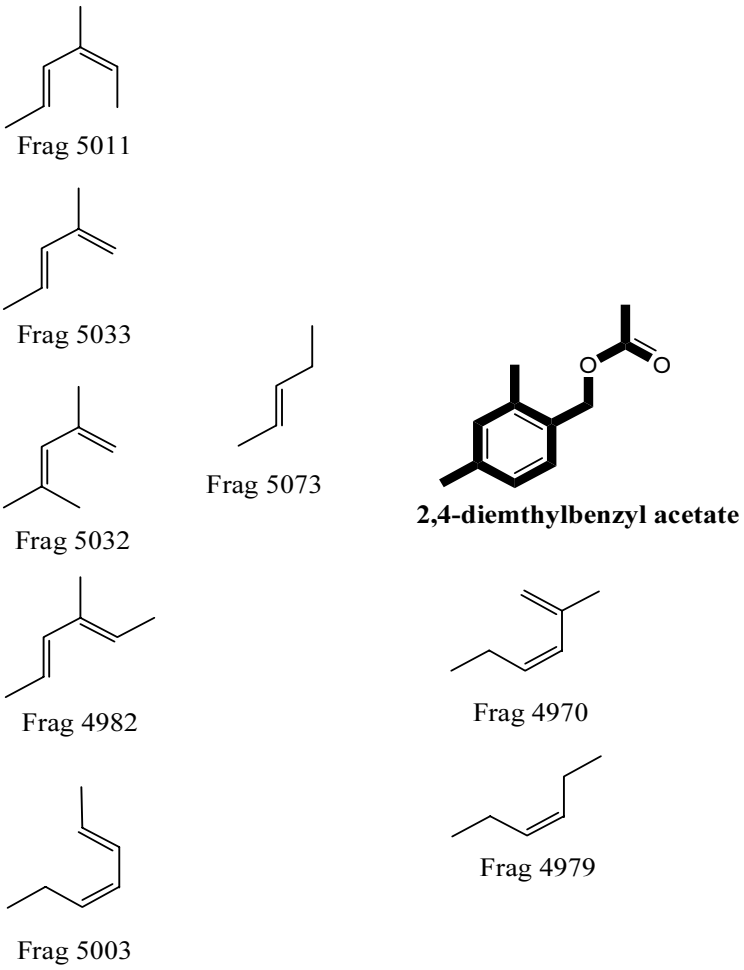

Figure 4.2 Illustration of the eight significant fragments contributing to the inactive validation prediction of the non-sensitizer 2,4-dimethylbenzyl acetate.

4.1.6 Identifying Structural Alerts

The chemical fragments illustrated in the preceding figures were chosen for two reasons: (1) it was feasible to select chemicals with few significant fragments as some compounds generated hundreds of key features contributing to its prediction. This made it more practical for clear illustration; and (2) by using the compounds whose mechanisms are already well documented and understood in the literature made it possible to verify that the models were mechanistically sound. Ashby et al established a group of structural alerts described on the basis of their inherent electrophilicity or electrophilic metabolites for DNA

reactivity and cancer (Ashby and Paton, 1993). These structural alerts for carcinogens, as described by Ashby, were used as a guide in explaining the cat-SAR system's basis for identifying mechanistically justifiable structural features of genotoxic and non-genotoxic carcinogens (i.e., aromatic amines, nitrogen mustards, epoxides).

It is important to note that the chemical predictions being made are solely dependent upon results of the LOO-CV procedure. As discussed, each removed chemical is predicted based on the significant fragments of the model's remaining chemicals. Hence, the chemicals own fragments are not contributive to its active or inactive prediction.

4.1.6.1 1-Phenyl-3,3-dimethyltriazene (PDMT)

1-Phenyl-3,3-dimethyltriazene (PDMT) was selected to demonstrate cat-SAR predictions of non-mammary carcinogens based on the rat MC-NMC model. PDMT is an alkylating agent with strong antimutagenic effects of fluoride on its mutation induction in *Drosophila melanogaster* (Vogel 1973). This compound is the most well known triazene used in anticancer studies. The formation of O6**-**methyldeoxyguanosine (O6-MedG) by PDMT in DNA and O6-alkylguanine-DNA alkyltransferase (ATase) in human peripheral leukocytes has been observed (Lee et al 1994). This compound has not been tested in mice. However, based on its classification in the CPDB, PDMT has been shown to induce cancer of the nervous system.

This *Salmonella* mutagen was correctly predicted by the cat-SAR program as an inactive compound in terms of its inability to induce mammary gland tumors in rats. PDMT was shown to have a 98% probability of being a non-mammary carcinogen (Table 4.14). Based upon the eight structural alerts shown to be responsible for the non-induction of mammary tumors in the rat by the cat-SAR program, PDMT's inactivity appear to be due in

large to the N-N=N chemical group extending from the phenyl ring (Figure 4.3). The depicted fragments (with the exception of fragment 1552) were all derived from other rat non-mammary carcinogens.

Table 4.14 Fragments from the ABC model leave-one-out validation analysis used to predict the inactivity of the rat non-mammary gland carcinogen 1-phenyl-3,3-dimethyltriazene.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag1552	1	11	12	0.083	0.917
Frag1557	0	5	5	0.000	1.000
Frag1558	0	5	5	0.000	1.000
Frag1559	0	5	5	0.000	1.000
Frag1561	0	5	5	0.000	1.000
Frag1572	0	6	6	0.000	1.000
Frag1576	0	6	6	0.000	1.000
Frag1577	0	5	5	0.000	1.000
	Probability of activity			0.02	0.98

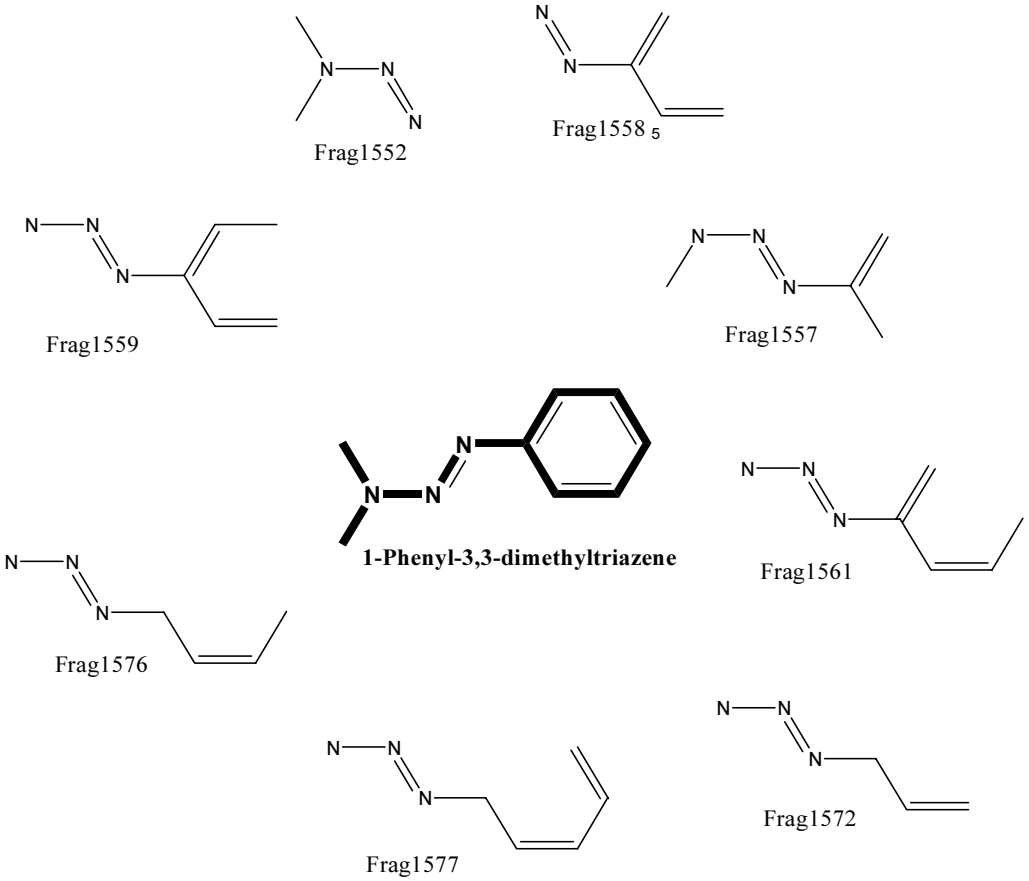

Figure 4.3 Illustration of the 8 significant fragments contributing to the inactive validation prediction of the rat non-mammary carcinogen 1-phenyl-3,3-dimethyltriazene.

4.1.6.2 Nithiazide

Nithiazide was selected to illustrate cat-SAR prediction of mammary carcinogens based on the rat MC-NMC model. This chemical agent is a synthetic antiprotozoal agent used in poultry farming. Though epidemiological studies have provided "sufficient" evidence of nithiazide's carcinogenicity in humans, the carcinogenicity tests in male and female rats and mice are considered to be inconclusive because of limitations in the design and results of these tests (IARC, 1981, 1987). Humans may be exposed as a result of its manufacture and use in veterinary medicine. Nithiazide has been tested for carcinogenicity in one experiment in mice and in one experiment in rats by administration in the diet. It increased the incidence of hepatocellular carcinomas and adenomas in male mice. Additionally, nithiazide does not induce mammary tumors in male rats. In female rats, it increased the incidences of fibroadenomas and cystadenomas of the skin, subcutaneous tissue and mammary gland and the incidence of endometrial stromal polyps of the uterus.

The 29 structural alerts responsible for the mammary carcinogenicity of the chemical nithiazide were identified (Figure 4.4). Of interest, the genotoxic nitro (NO_2) component was excluded from the depiction of structural features identified to be responsible for the induction of mammary tumorigenesis. Although, this chemical group is considered a structural alert for carcinogenicity (mutagenicity) when examining cancer in the general rodent data it is not significant according to this study in terms of breast cancer induction. The analyses suggest that this genotoxic chemical elicits breast carcinogenic effects due to the presence of the molecular fragments shown in Figure 4.2. In particular, fragments of the C2-S-C ring structure and S-C-N group appear to contribute heavily to the breast

carcinogenic action of nithiazide. The finding was based on the presence of these groups in several other mammary gland carcinogens in the model.

There are several reasons to possibly assess the exclusion of the nitro ($-NO_2^-$) component based on the findings of the cat-SAR program: (1) the nitro component may not be associated with mammary carcinogenicity, (2) this structural feature may not discriminate carcinogens from non-carcinogens, and/or (3) this fragment may have not met the set criteria and as a result was eliminated from the final model. More than likely, it is speculated that the third reason may be the most concise. It is possible that this genotoxic structural alert to carcinogenicity was found in both general carcinogens and breast carcinogens. In other words, the nitro component was present on both sides of the MC-NMC model. Thus, it would be ruled out based on the user criterion governing structural features found equally in active and inactive compounds. With respect to the first possible explanation, it is very unlikely that the NO_2 electrophile is not a structural alert to mammary cancer considering a lot of nitro-containing chemicals have been identified to be mammary carcinogens (i.e., contain the genotoxic nitro group that binds to macromolecular structures such as DNA). Studies have been conducted that demonstrate nitro-mediated mammary tumorigenesis (Jadeski et al 2003). Thus, providing reason to believe that nitro activity is positively associated with breast cancer progression. Jadeski *et al.* demonstrated specifically how chemicals that serve as nitro donors stimulate phosphorylation of extracellular signal-regulated kinases (ERK), demonstrating a role for endogenous and exogenous nitro groups in ERK activation (Jadeski et al 2003).

The genotoxic chemical nithiazide was predicted to be a rat mammary gland carcinogen based on 29 structural alerts found predominately in mammary carcinogens

(Figure 4.4). Six of the 29 fragments also included some inactive components. This contributed to the compound having a 97.7% probability of being a rat mammary gland carcinogen (Table 4.15).

Table 4.15 Fragments from the ABC model leave-one-out validation analysis used to predict the activity of the rat mammary gland carcinogen nithiazide.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag328	11	1	12	0.917	0.083
Frag352	11	1	12	0.917	0.083
Frag361	12	1	13	0.923	0.077
Frag508	21	2	23	0.913	0.087
Frag746	12	1	13	0.923	0.077
Frag756	12	1	13	0.923	0.077
Frag1739	10	0	10	1.000	0.000
Frag1740	10	0	10	1.000	0.000
Frag1741	10	0	10	1.000	0.000
Frag1742	9	0	9	1.000	0.000
Frag1743	9	0	9	1.000	0.000
Frag1744	9	0	9	1.000	0.000
Frag1745	9	0	9	1.000	0.000
Frag1746	9	0	9	1.000	0.000
Frag1747	9	0	9	1.000	0.000
Frag1754	9	0	9	1.000	0.000
Frag1758	10	0	10	1.000	0.000
Frag1759	9	0	9	1.000	0.000
Frag1765	10	0	10	1.000	0.000
Frag1775	11	0	11	1.000	0.000
Frag1776	11	0	11	1.000	0.000
Frag1777	10	0	10	1.000	0.000
Frag1778	10	0	10	1.000	0.000
Frag1779	10	0	10	1.000	0.000
Frag1780	10	0	10	1.000	0.000
Frag1781	10	0	10	1.000	0.000
Frag1782	10	0	10	1.000	0.000
Frag1801	10	0	10	1.000	0.000
Frag1831	10	0	10	1.000	0.000
	Probability of activity			0.977	0.023

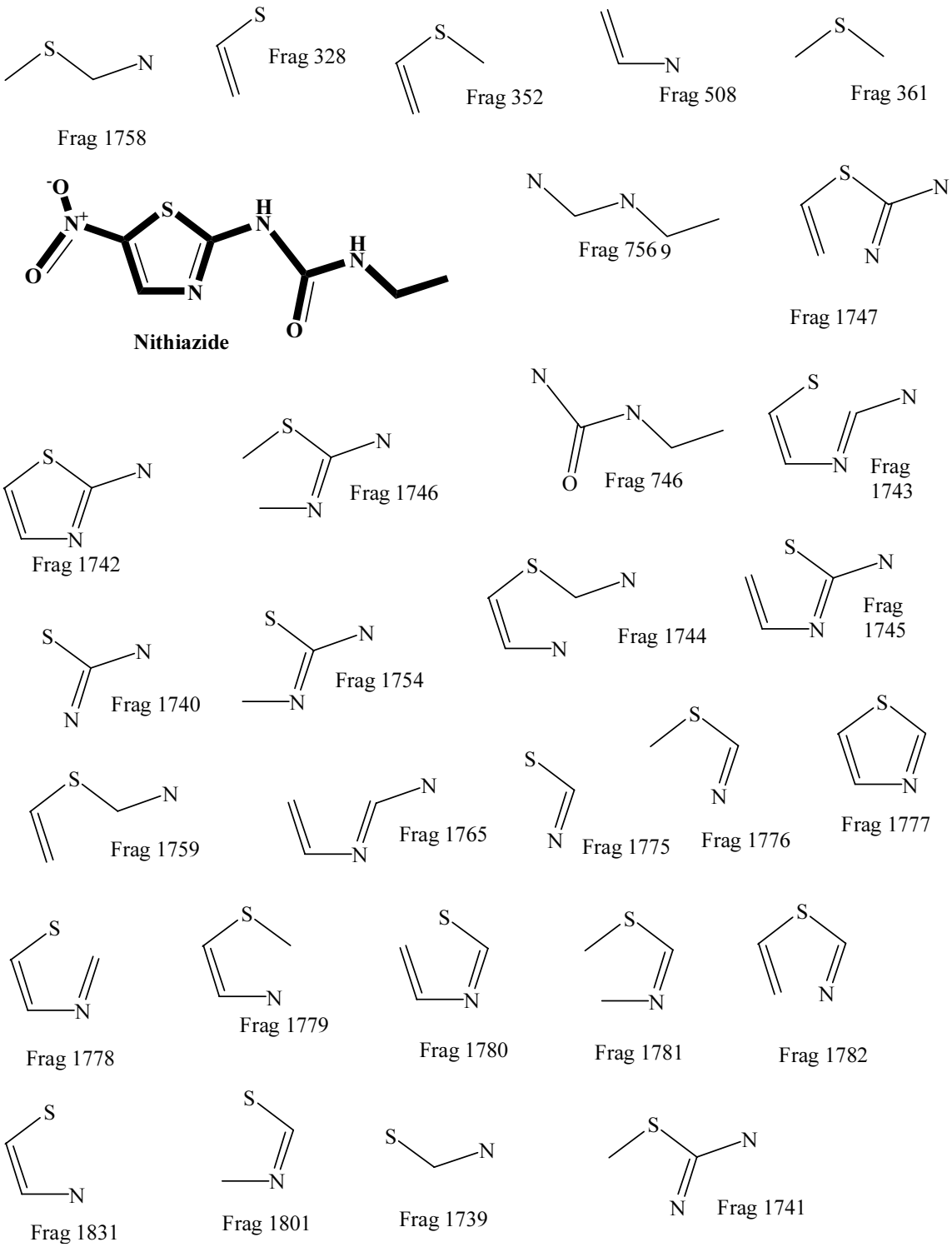

Figure 4.4 Illustration of the 29 significant fragments contributing to the active validation prediction of the rat mammary carcinogen nithiazide.

4.1.6.3 Fenaminosulf

Formulated fenaminosulf was selected to illustrate the cat-SAR program's utility in distinguishing rat noncarcinogens from mammary carcinogens based on the rat MC-NC model. Fenaminosulf (p-dimethylaminobenzenediazo sodium sulfonate), an active ingredient in several commercial fungicides, was reported to be mutagenic in *Salmonella typhimurium* (McCann et al 1975). Since fenaminosulf has structural similarity to the potent carcinogen, butter yellow (p-dimethylaminoazobenzene), it has been evaluated for possible mutagenicity in *Drosphila melanogaster* (Pai 1983) and also for its carcinogenic potential in the rodent bioassay (NTP 1978). No statistically significant positive associations were demonstrated. There are conflicting reports concerning this compound's ability to induce hepatomas in rats. However, based on this aromatic diazo compound's classification in the CPDB, fenaminosulf does not induce cancer in male and female rats and mice.

The cat-SAR program identified seven fragments responsible for the deactivation of formulated fenaminosulf in rats resulting in an inactive prediction being made for the compound (Table 4.15). Each of the seven fragments was observed in a total of four other compounds in the dataset. All four of these chemicals were non-mammary carcinogens.

Table 4.16 Fragments from the ABC 3/0.90 model leave-one-out validation analysis used to predict the inactivity of the rat noncarcinogen fenaminosulf.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag 6443	0	4	4	0.000	1.000
Frag 6446	0	4	4	0.000	1.000
Frag 6447	0	4	4	0.000	1.000
Frag 6451	0	4	4	0.000	1.000
Frag 6452	0	4	4	0.000	1.000
Frag 6455	0	4	4	0.000	1.000
Frag 6461	0	4	4	0.000	1.000
	Probability of activity			0.00	1.00

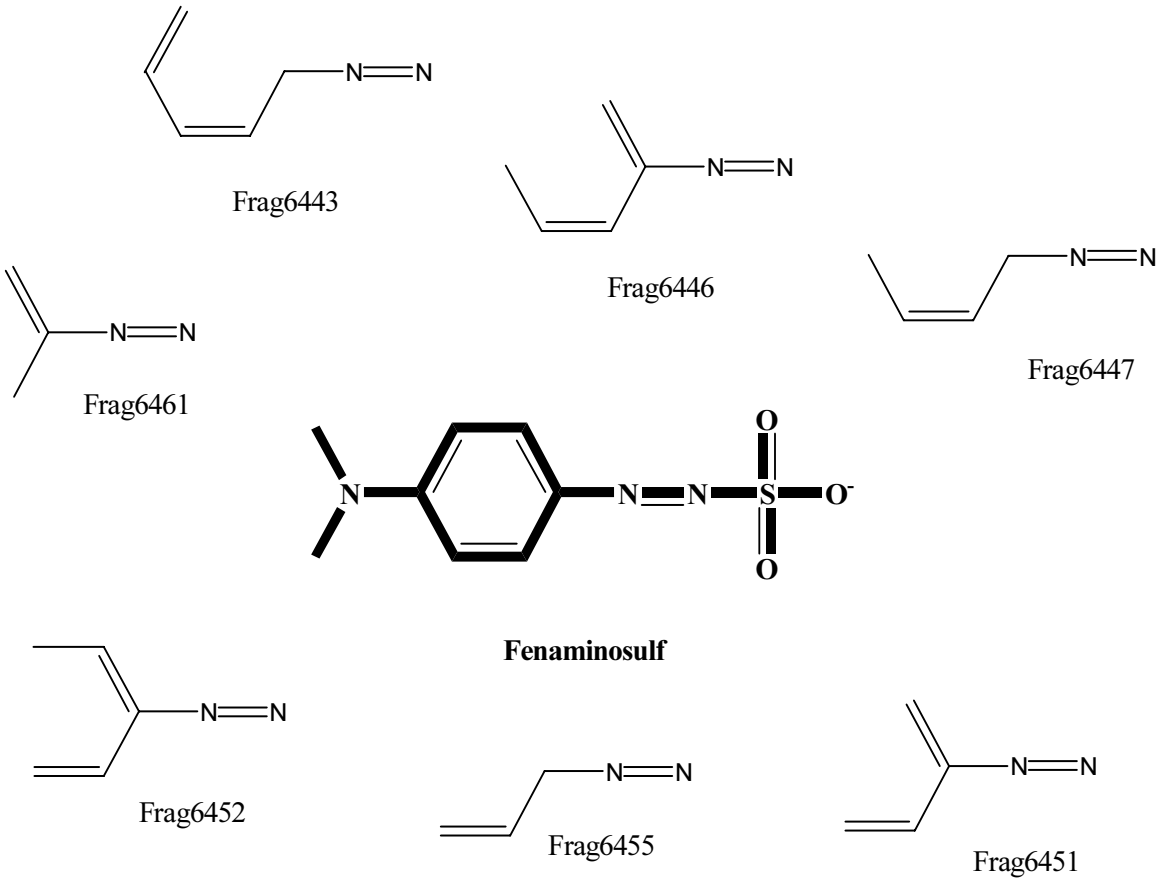

Figure 4.5 Illustration of the 7 significant fragments contributing to the inactive validation prediction of the rat noncarcinogen fenaminosulf.

4.1.6.4 Atrazine

Atrazine (ATR) was selected to illustrate cat-SAR's potential to predict mammary carcinogens based on the rat MC-NC model. Atrazine was one of the very few male-only mammary carcinogens in the MC-NC learning set. ATR does not induce tumors at any other site in the male rat. According to atrazine's classification in the CPDB, ATR has been shown to induce tumors in the hematopoietic system and also the uterus in female rats. Hazard assessment of the widely used chloro-S-triazine herbicide, atrazine, has largely focused on the compound's induction of mammary tumors. ATR is not estrogenic and most studies have

found that it is not mutagenic in *Salmonella typhimurium* (Brusick 1994). Atrazine has been tested for mutagenic potential in more than 50 studies of gene mutation, chromosomal aberration, and primary DNA damage, and a weight-of-evidence evaluation indicates a nonmutagenic status relative to conventional health effects-testing formats (Brusick 1994).

Atrazine is metabolized in mammals principally by dealkylation of the amino groups (Eldridge et al 1994). Studies in humans have shown that metabolism proceeds in much the same manner as in rodents (Adams et al 1990), However, one might surmise that the rat model is more sensitive than other species, including humans, to hormone-related effects because metabolism and clearance in rats proceed at a slower rate (Eldridge et al 1994).

ATR was correctly predicted to be a rat mammary carcinogen based on ten fragments (Figure 4.6). All ten structural alerts identified were also present in three other mammary carcinogens. ATR was predicted to have a 100% probability of being a mammary carcinogen in rats (Table 4.16).

Table 4.17 Fragments from the ABC 3/0.90 model leave-one-out validation analysis used to predict the activity of the rat mammary gland carcinogen atrazine.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag 3662	3	0	3	1.000	0.000
Frag 3663	3	0	3	1.000	0.000
Frag 3664	3	0	3	1.000	0.000
Frag 3665	3	0	3	1.000	0.000
Frag 3666	3	0	3	1.000	0.000
Frag 3667	3	0	3	1.000	0.000
Frag 3668	3	0	3	1.000	0.000
Frag 3670	3	0	3	1.000	0.000
Frag 3671	3	0	3	1.000	0.000
Frag 3677	3	0	3	1.000	0.000
	Probability of activity			1.000	0.000

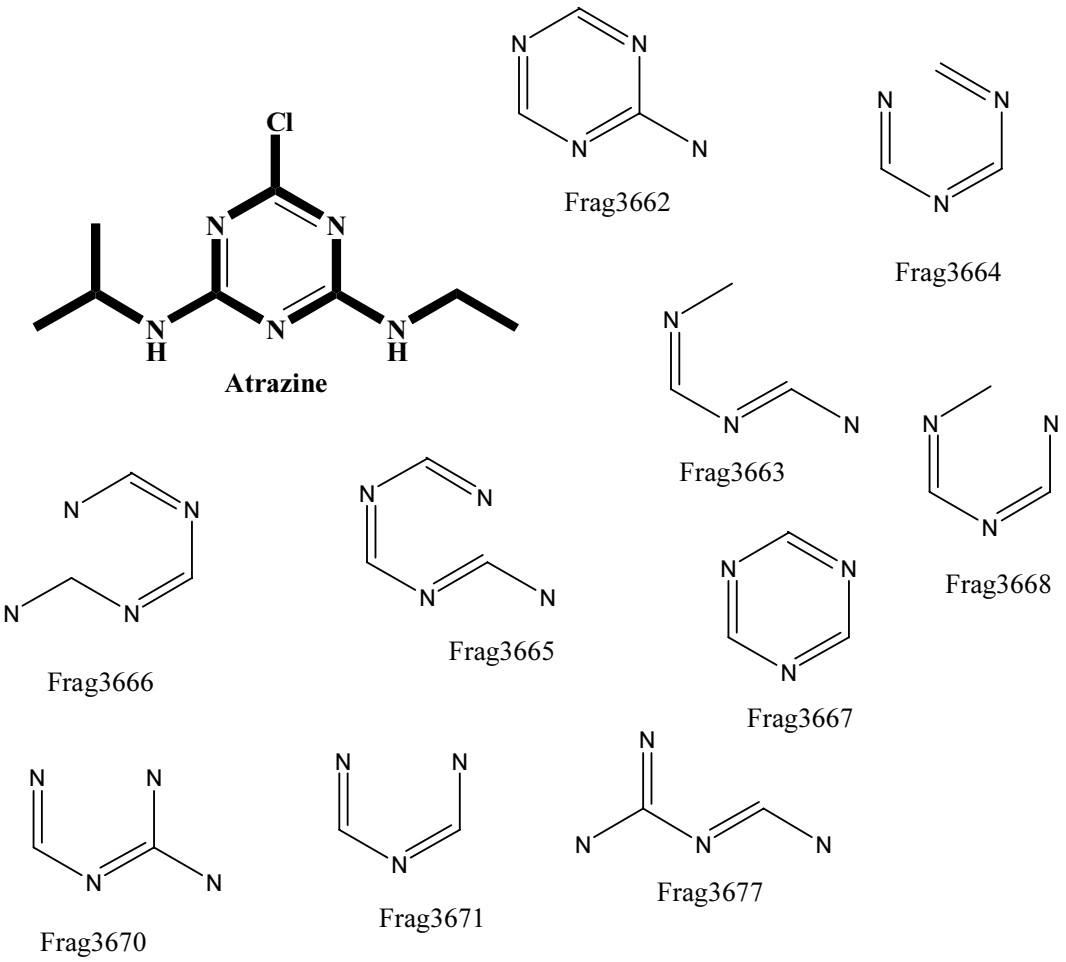

Figure 4.6 Illustration of the 10 significant fragments contributing to the active validation prediction of the rat mammary carcinogen atrazine.

4.1.6.5 Diazepam

Diazepam was selected to illustrate cat-SAR prediction of mouse non-mammary carcinogens based on the mouse MC-RNC model. Epidemiological studies have found no positive relation with breast cancer risk or with the extent of disease and lymph node involvement, and the possibility of a protective effect has been suggested (Kleinerman et al 1984). Additionally, diazepam has been evaluated in *Salmonella* and found to be non-mutagenic.

Diazepam was predicted by cat-SAR to be a mouse non-mammary carcinogen. Diazepam has been tested in both male and female rats and mice and has not been observed

to induce cancer in any of the four groups. Each of the eight fragments were predominately found in other non-mammary carcinogens (Table 4.17 and Figure 4.7). However, 4 of the eight fragments included inactive features. As such, diazepam was predicted to have an 84% probability of being a mouse non-mammary carcinogen.

Table 4.18 Fragments from the ABC 3/0.75 model leave-one-out validation analysis used to predict the inactivity of the rodent non-carcinogen diazepam.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag188	3	9	12	0.250	0.750
Frag201	1	6	7	0.143	0.857
Frag238	2	6	8	0.250	0.750
Frag1340	1	3	4	0.250	0.750
Frag1346	0	3	3	0.000	1.000
Frag1351	0	3	3	0.000	1.000
Frag1385	0	3	3	0.000	1.000
Frag1397	0	3	3	0.000	1.000
	Probability of activity			0.163	0.837

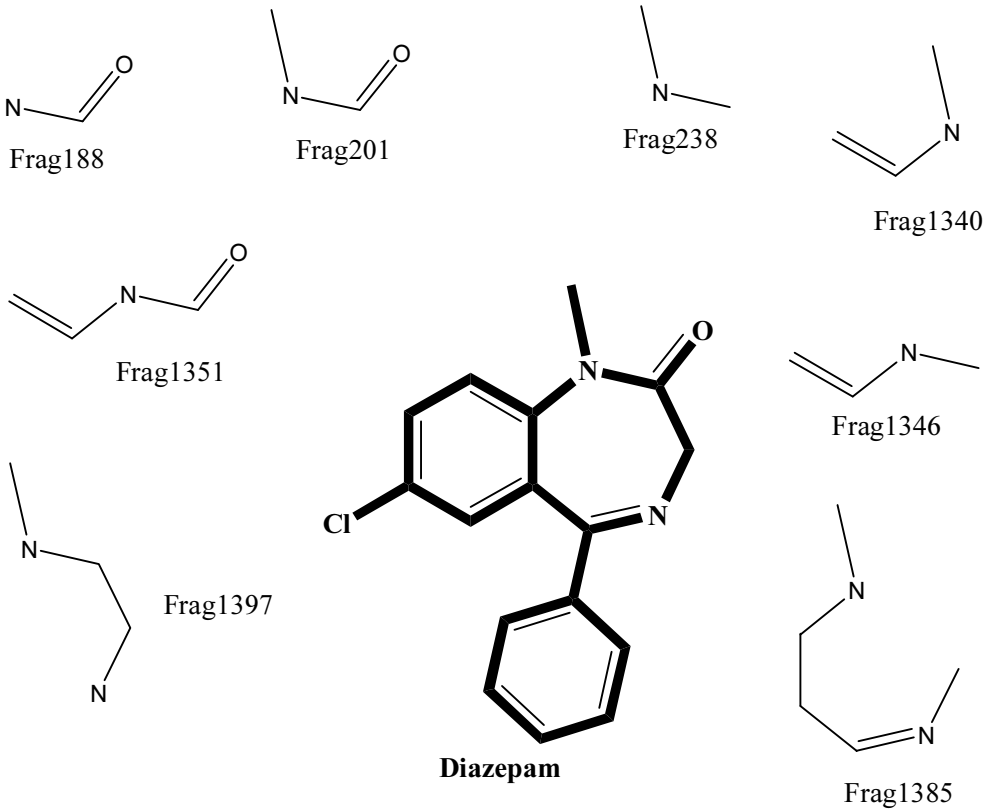

Figure 4.7 Illustration of the 8 significant fragments contributing to the inactive validation prediction of the rodent non-carcinogen diazepam.

4.1.6.6 Calciferol

Calciferol (i.e., vitamin D) was selected to illustrate cat-SAR prediction of mouse mammary carcinogens based on the mouse MC-RNC model. Calciferol has only been tested in mice and has not been shown to induce tumors at any other site. Calciferol was correctly predicted to be a mouse mammary carcinogen based on the possession of 23 significant fragments (Figure 4.8). All 23 fragments were predominately found in mammary carcinogens. However, 7 of the 23 structural alerts also included some inactive features (Table 4.18). As such, calciferol was predicted to have a 92.4% probability of being a mouse mammary carcinogen. There is no *Salmonella* mutagenicity data for this compound.

Table 4.19 Fragments from the ABC 3/0.75 model leave-one-out validation analysis used to predict the activity of the mouse mammary gland carcinogen calciferol.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag9	4	0	4	1.000	0.000
Frag333	8	1	9	0.889	0.111
Frag372	7	1	8	0.875	0.125
Frag382	6	0	6	1.000	0.000
Frag383	5	0	5	1.000	0.000
Frag388	4	1	5	0.800	0.200
Frag638	3	0	3	1.000	0.000
Frag640	3	0	3	1.000	0.000
Frag641	3	0	3	1.000	0.000
Frag642	3	0	3	1.000	0.000
Frag652	3	1	4	0.750	0.250
Frag653	3	0	3	1.000	0.000
Frag679	3	0	3	1.000	0.000
Frag680	3	0	3	1.000	0.000
Frag682	3	0	3	1.000	0.000
Frag685	3	0	3	1.000	0.000
Frag702	3	0	3	1.000	0.000
Frag750	3	1	4	0.750	0.250
Frag796	3	1	4	0.750	0.250
Frag846	3	1	4	0.750	0.250
Frag938	3	0	3	1.000	0.000
Frag946	3	0	3	1.000	0.000
Frag947	3	0	3	1.000	0.000
	Probability of activity			0.924	0.076

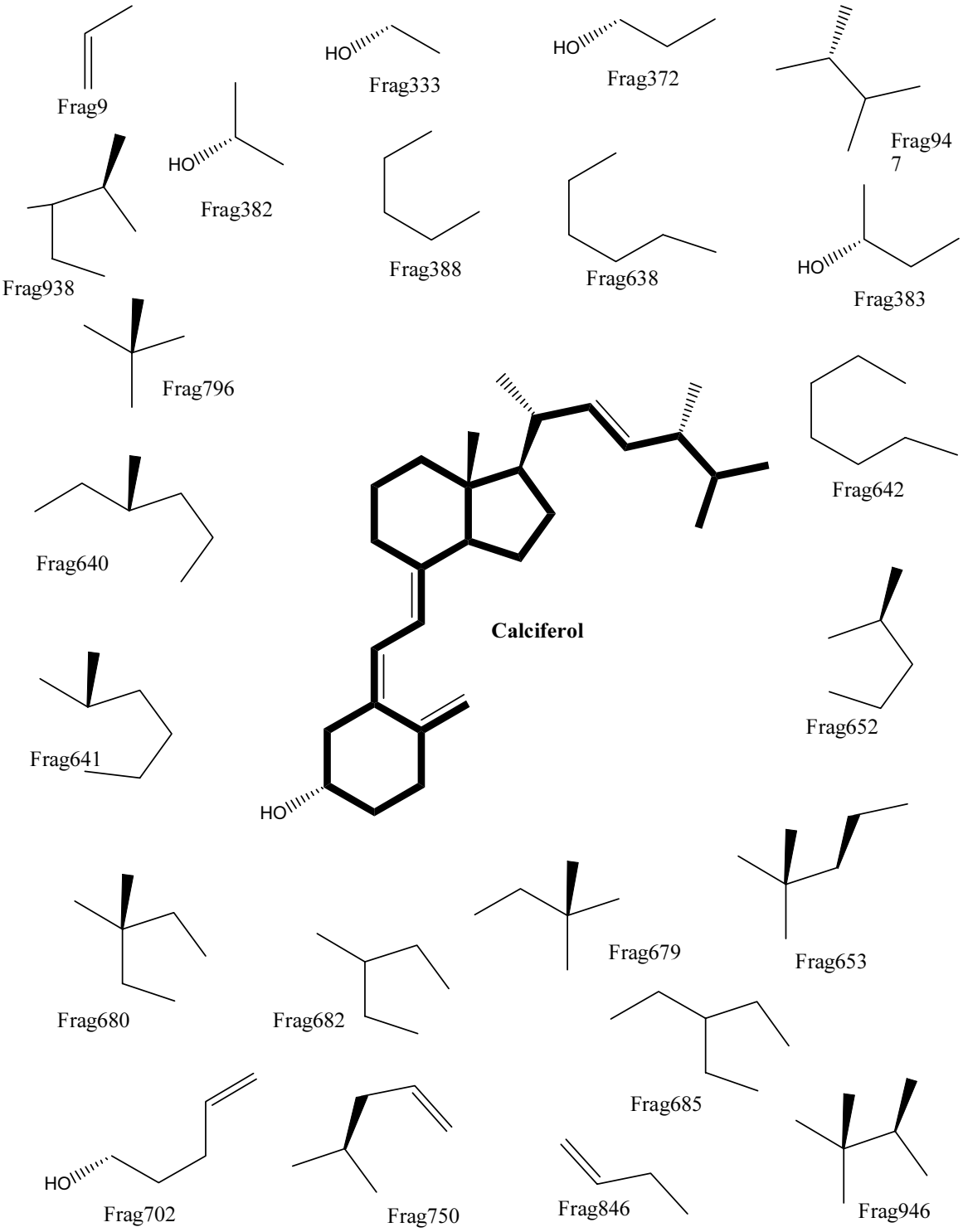

Figure 4.8 Illustration of the 23 significant fragments contributing to the active validation prediction of the mouse mammary carcinogen calciferol.

4.1.6.7 1'-Hydroxysafrole

1'-Hydroxysafrole was selected to illustrate the cat-SAR program's prediction of mammary carcinogens based on the mouse MC-NMC model. Based on its classification in the CPDB, 1'-hydroxysafrole has been tested in male and female mice and male rats. 1'-Hydroxysafrole is a proximate carcinogenic metabolite of the naturally occurring hepatocarcinogen safrole. The sulfuric acid ester of 1'-hydroxysafrole, namely 1'-sulfoxysafrole, was shown to be the major ultimate electrophilic and carcinogenic metabolite of 1'-hydroxysafrole in mouse liver (Miller et al 1983). The sulfuric acid ester of 3'-hydroxyisosafrole was also shown to be electrophilic when chemically synthesized or generated enzymatically *in vitro* (Miller et al 1983). However, no evidence has been obtained for *in vivo* metabolism of 3'-hydroxyisosafrole to a sulfuric acid ester in the mouse.

In one study, administration of 1'-hydroxysafrole to mice in the diet for 4-14 days caused a 90% inhibition of the covalent binding of a subsequent dose of 1'-hydroxysafrole to the hepatic macromolecules (Miller 1983). This effect was due to increased detoxification of 1'-hydroxysafrole rather than to an inhibition of metabolic activation. Under the test conditions, 1'-hydroxysafrole had very little, if any, tumor-initiating activity in rat liver, but did exhibit strong promoting activity. This promoting activity was inhibited almost completely by pentachlorophenol, indicating that it was mediated by the electrophilic sulfuric acid ester of 1'-hydroxysafrole (Miller et al 1983).

Sulfation is a common final step in the biotransformation of xenobiotics and is traditionally associated with inactivation. However, the sulfate group is electron withdrawing and may be cleaved off heterolytically in some molecules leading to electrophilic cations which may form adducts with DNA and other important cellular structures (Glatt 1998).

Since endogenous sulfotransferases do not appear to be expressed in indicator cells of standard mutagenicity tests, rat and human sulfotransferases have been stably expressed in his-*Salmonella typhimurium* strain TA1538 and Chinese hamster V79 cells by Glatt and collaborators. Using these recombinant indicator cells, sulfotransferase-dependent genotoxic activities were detected with 1'-hydroxysafrole (Glatt 1998). In other cases, spontaneous benzylic substitution reactions with medium components, such as halogenide ions or amino acids, led to secondary, membrane-penetrating reactive species.

Different sulfotransferases, including related forms from rat and human, substantially differed in their substrate specificity towards the investigated promutagens. It is known that some sulfotransferases are expressed with high tissue and cell type specificities (Glatt 1998). As described by Glatt, this site-dependent expression together with the limitations in the distribution of reactive sulfuric acid conjugates may explain organotropic effects of compounds activated by this metabolic pathway (Glatt 1998).

Herein, the cat-SAR program correctly classified 1'-hydroxysafrole as a non-mammary carcinogen based on 12 structural alerts associated with mammary tumorigenesis in other compounds in the mouse MC-NMC learning set (Figure 4.9). Based on these findings, it was noted that the inactivity of the non-mammary carcinogen, 1'-hydroxysafrole, is due in large to the compound's benzyl ring. Furthermore, it was noted that the oxygen-containing parts of the molecule were excluded as significant inactive fragments. It is more than likely that the electron-withdrawing effect of the five-membered ring contributes to the 23% chance of activity. It was noted that all 12 fragments had included mostly 1 or 2 active components. As a result, 1'hydroxysafrole was predicted to have a 77% probability of being a mouse non-mammary gland carcinogen.

Table 4.20 Fragments from the ABC 3/0.75 model leave-one-out validation analysis used to predict the inactivity of the mouse non-mammary gland carcinogen 1'-hydroxysafrole.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
79	1	5	6	0.167	0.833
80	1	5	6	0.167	0.833
81	1	5	6	0.167	0.833
82	1	5	6	0.167	0.833
83	1	5	6	0.167	0.833
86	1	6	7	0.143	0.857
89	2	6	8	0.250	0.750
110	1	6	7	0.143	0.857
111	1	5	6	0.167	0.833
118	2	6	8	0.250	0.750
126	1	6	7	0.143	0.857
144	5	1	6	0.833	0.167
	Probability of activity			0.228	0.772

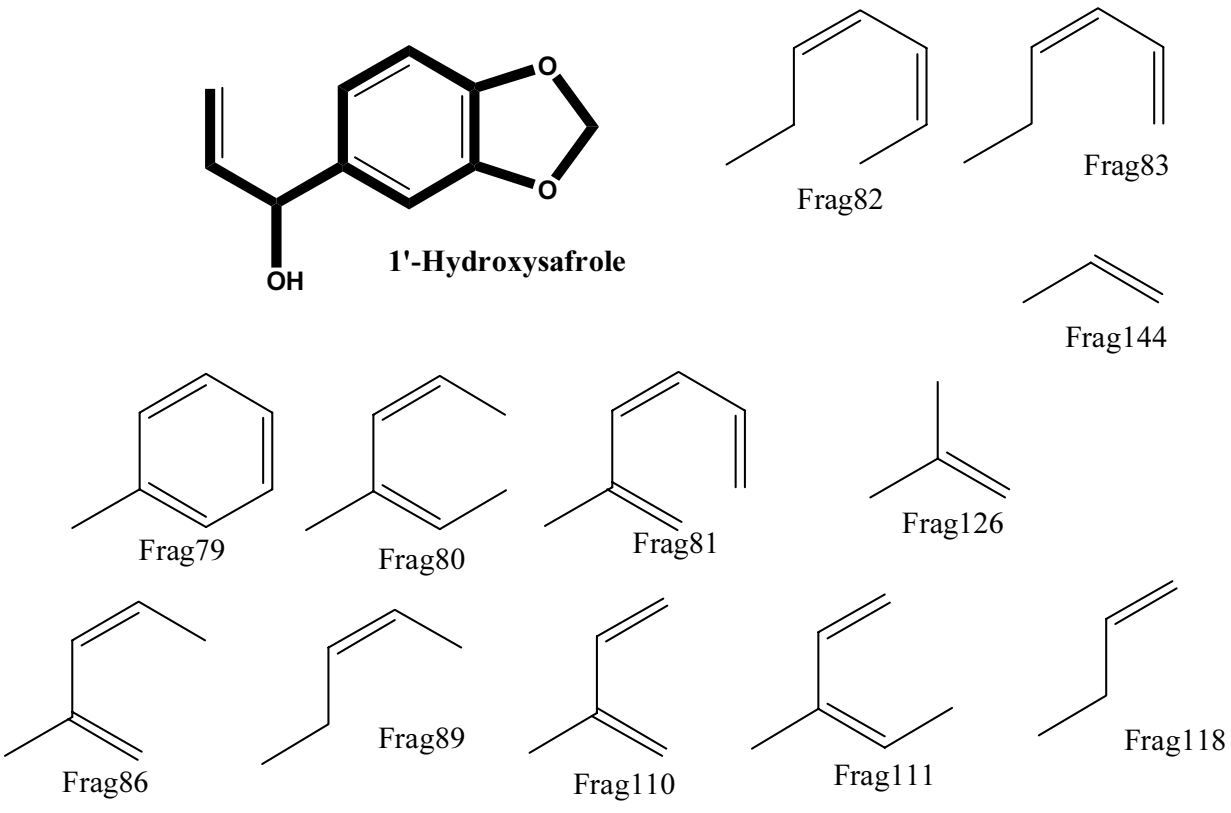

Figure 4.9 Illustration of the 12 significant fragments contributing to the inactive validation prediction of the mouse non-mammary gland carcinogen 1'-hydroxysafrole.

4.1.6.8 Diethylstilbestrol

The synthetic estrogen, diethylstilbestrol (DES) is a drug that was prescribed to pregnant women between the late 1930's to early 1970's to prevent miscarriages or early termination of pregnancy. However, physicians were advised by the Food and Drug Administration to discontinue its prescription due to its link to a rare from of vaginal cancer called clear cell adenocarcinoma. DES is a known human carcinogen (i.e., transplacental cancer) in the daughters of women that were exposed to this drug. DES is well documented to be a perinatal carcinogen in both humans and experimental animals. As described by Heneweer *et al.*, "approximately 60% of all breast tumors are estrogen responsive and chemicals that show estrogenic or anti-estrogenic properties are able to interact with breast tumor growth. In a breast tumor, adipose stromal cells (fibroblasts) surrounding the epithelial tumor contain the aromatase enzyme, which converts androgens into estrogens." (Heneweer et al 2004). It is this exposure to aromatase inducers that can then result in increased estrogen levels, which potentially can increase breast tumor growth.

DES was selected to illustrate cat-SAR's prediction of mouse mammary carcinogens based on the mouse MC-NMC model. DES has been tested in male and female mice and rats and has been observed to induce cancer in all four groups. This compound induces mammary tumors and cancer of the thyroid and pituitary gland in both male and female mice. However, DES is not classified as a *Salmonella* mutagen. Based on this compound's classification in the CPDB, DES induces cancer of the adrenal and pituitary gland in male rats and the liver and pituitary gland in female rats. Additionally, DES induces testicular cancer in male mice and uterine and ovarian cancer in female mice.

Diethylstilbestrol was predicted to be a mouse mammary carcinogen based on its possession of 17 structural alerts (Figure 4.10). Based on these findings, the presence of the two phenolic hydroxyl groups seem to be responsible for the toxicological activity of the compound (i.e., a "neighboring effect"). Each of the seventeen fragments was found in three to five other compounds in the learning set, all of which induced mammary tumorigenesis. However, fragment 1115 also included 1 inactive component. Therefore, diethylstilbestrol was predicted to have a 98.6% probability of being a mouse mammary gland carcinogen (Table 4.21).

Table 4.21 Fragments from the ABC 3/0.75 model leave-one-out validation analysis used to predict the activity of the mouse mammary gland carcinogen diethylstilbestrol (DES).

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag1115	4	1	5	0.800	0.200
Frag1223	3	0	3	1.000	0.000
Frag2002	3	0	3	1.000	0.000
Frag2003	4	0	4	1.000	0.000
Frag2004	4	0	4	1.000	0.000
Frag2005	4	0	4	1.000	0.000
Frag2006	4	0	4	1.000	0.000
Frag2007	4	0	4	1.000	0.000
Frag2010	3	0	3	1.000	0.000
Frag2011	3	0	3	1.000	0.000
Frag2014	4	0	4	1.000	0.000
Frag2017	4	0	4	1.000	0.000
Frag2024	5	0	5	1.000	0.000
Frag2025	5	0	5	1.000	0.000
Frag2034	5	0	5	1.000	0.000
Frag2047	5	0	5	1.000	0.000
Frag2060	5	0	5	1.000	0.000
	Probability of activity			0.986	0.014

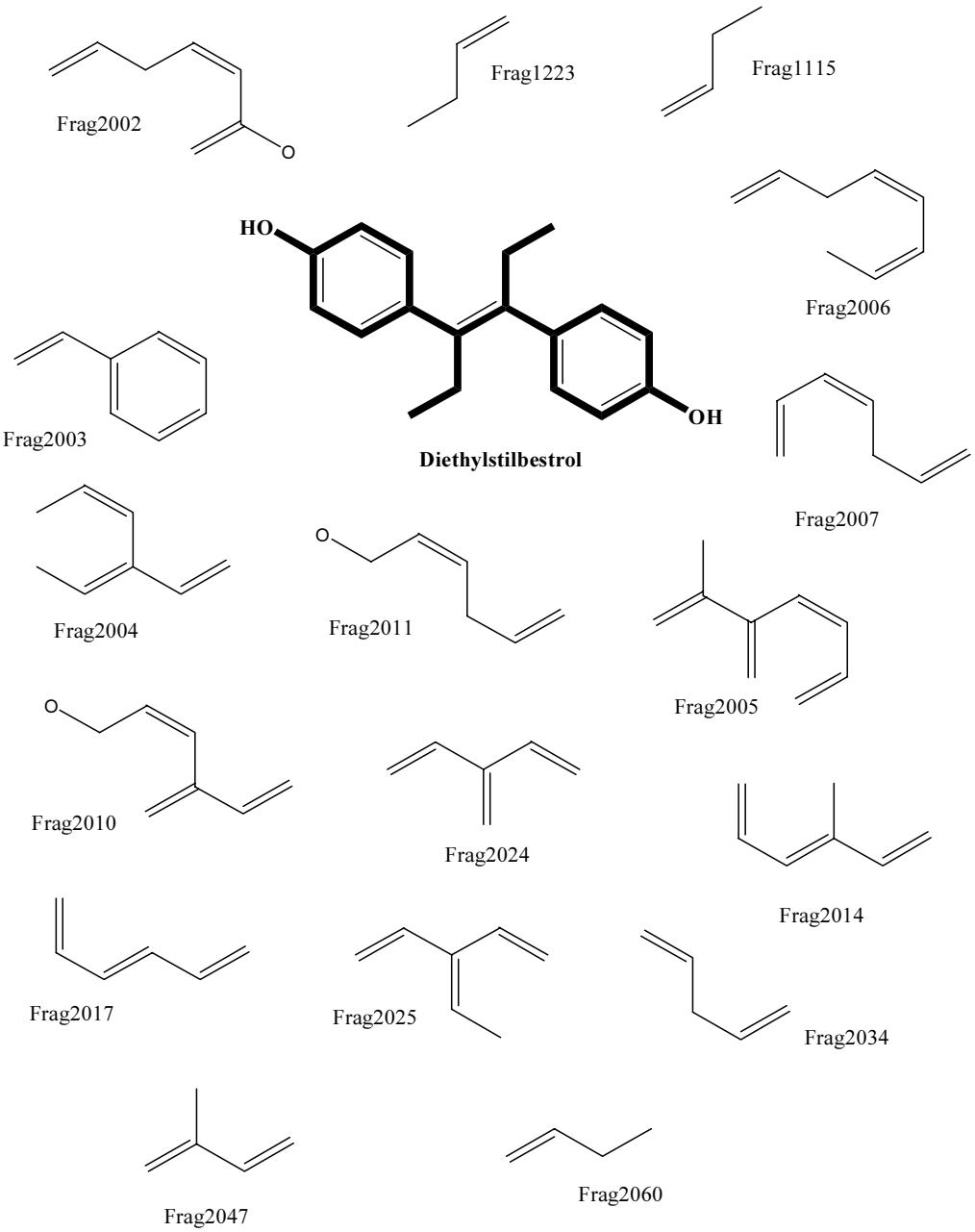

Figure 4.10 Illustration of the 17 significant fragments contributing to the active validation prediction of the mouse mammary carcinogen diethylstilbestrol.

4.1.6.9 Capsaicin

Capsaicin was selected to illustrate the cat-SAR predictions of mouse carcinogens in the CPDB general mouse model. Capsaicin (8-methyl-N-vanillyl-6-nonenamide) is a quinone that has been shown to regulate a wide variety of activities that require NF-kappa B activation (Singh et al 1996). This is important because viral replication, immune regulation,

and induction of various inflammatory and growth-regulatory genes require activation of a nuclear transcription factor (NF)-kappa B. Agents, such as capsaicin, that can block NF-kappa B activation have potential to block downstream responses mediated through this transcription factor (Singh et al 1996). There is evidence that capsaicin may have carcinogenic potential in humans and in biological systems designed to model human cancer (Azizan and Blevins 1995). Both positive and negative effects have been found in classical genetic toxicology assays with capsaicin. However, the capsaicin tested in most studies has been derived from pepper plant extracts, which is likely to display varying degrees of purity and possibly diverse impurity profiles (Chanda et al 2004).

Capsaicin was predicted to be a mouse carcinogen based on the possession of 10 structural alerts associated with mouse carcinogenesis in other compounds in the learning set (Figure 4.11). Each of these fragments was present in three other compounds in the dataset, all of which were mouse carcinogens (Table 4.22). Furthermore, all structural alerts were derived from the benzyl ring of the compound's structure suggesting that aromaticity may play a role in the carcinogenicity of this compound.

Table 4.22 Fragments from the ABC 3/0.90 model leave-one-out validation analysis used to predict the activity of the mouse carcinogen capsaicin.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag7465	3	0	3	0.800	0.200
Frag7466	3	0	3	1.000	0.000
Frag7467	3	0	3	1.000	0.000
Frag7469	3	0	3	1.000	0.000
Frag7470	3	0	3	1.000	0.000
Frag7474	3	0	3	1.000	0.000
Frag7475	3	0	3	1.000	0.000
Frag7476	3	0	3	1.000	0.000
Frag7477	3	0	3	1.000	0.000
Frag7480	3	0	3	1.000	0.000
	Probability of activity			1.000	0.000

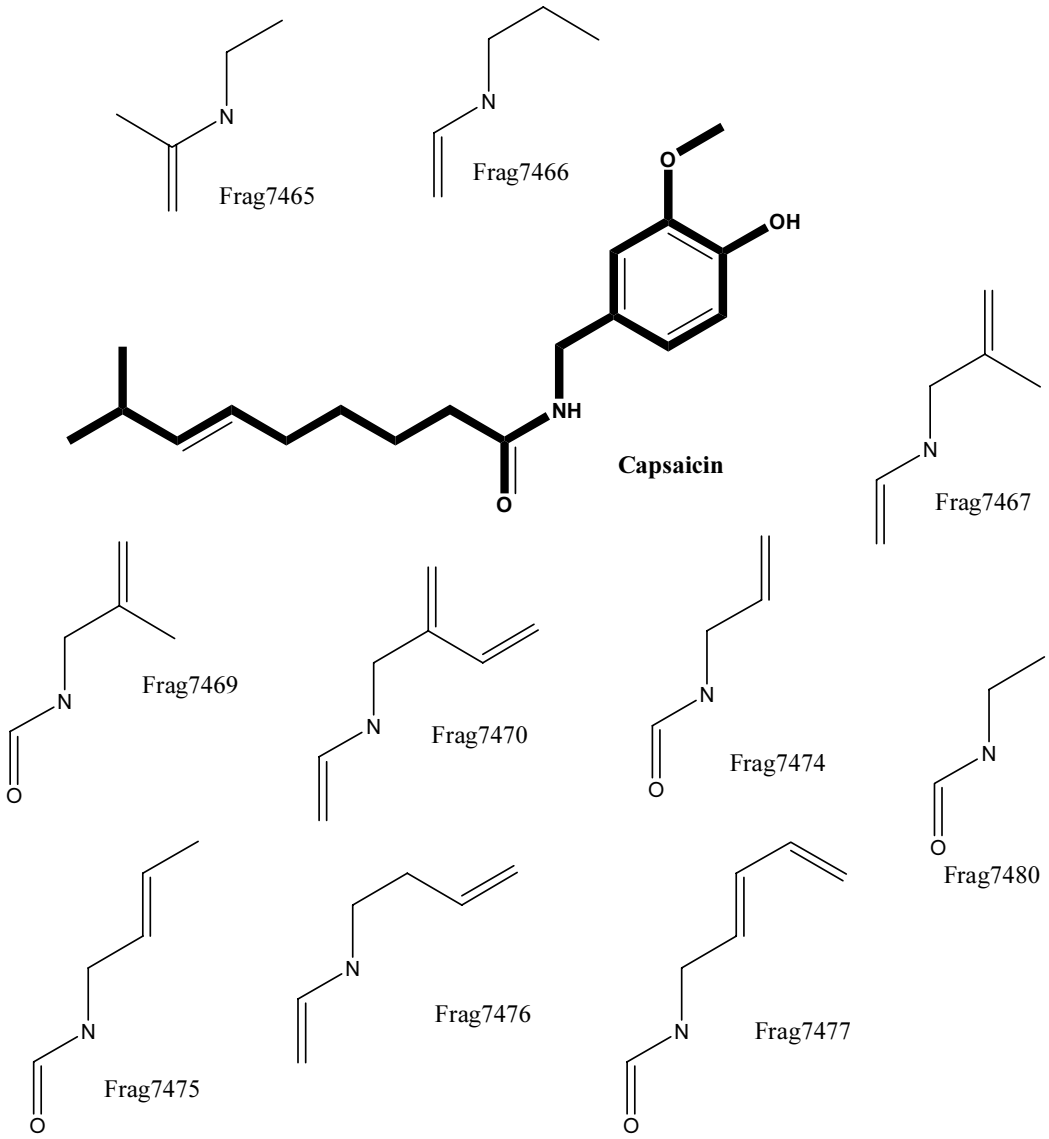

Figure 4.11 Illustration of the 10 significant fragments contributing to the active validation prediction of the mouse carcinogen capsaicin.

4.1.6.10 1-Trans-delta-9-tetrahydrocannabinol (Δ9-THC)

1-Trans-Δ9-tetrahydrocannabinol (Δ9-THC) was selected to illustrate cat-SAR predictions of mouse carcinogens based on the mouse carcinogenesis model. According to this compound's classification in the CPDB, Δ9-THC is not a *Salmonella* mutagen. This compound has been tested in male and female rats and mice and has not been observed to induce cancer in any of these groups.

Based on findings from the cat-SAR study, nucleophilicity may play a major role in the non-carcinogenic effect of this major component of marijuana. Other studies demonstrated the importance of the A-ring aryl C-3 side chain and phenolic hydroxyl substituents, and elucidated the importance of a C-ring hydroxyalkyl substituent in cannabinoid toxicity (Melvin et al 1993). Melvin et al examined the structure-activity relationship surrounding this region (D-ring) of the molecule that is not present in the structure of Δ9-THC and other classical cannabinoid compounds. Both rigid fused ring benzo and cyclohexyl derivatives (i.e., creating the D-ring) retained binding affinity for the cannabinoid receptor. Extension of ketone or hydroxyl substituents from the C-2 position of the D-ring resulted in a 3-fold increase in binding affinity over the unsubstituted structure. However, the fused ring structure is not critical for the interaction with the receptor in as much as opening the ring did not decrease the potency. Extension of the D-ring C-2 alcohol by one carbon in length resulted in a pair of structures, for which the greatest affinity for the CB1 receptor occurred for the hydroxymethyl group in the axial conformation.

The National Toxicology Program (NTP) performed carcinogenesis studies of Δ9-THC in rats and mice and concluded that under the conditions of their 2-year gavage studies, there was no evidence of carcinogenic activity of Δ9-THC in male or female rats (NTP 1996). However, there was equivocal evidence of carcinogenic activity of Δ9-THC in male and female mice based on the increased incidences of thyroid gland follicular cell adenomas when administered at 125mg/kg (NTP 1996). Increased incidences of thyroid gland follicular cell hyperplasia occurred in male and female mice, and increased incidences of hyperplasia and ulcers of the forestomach were observed in male mice (NTP 1996).

Δ9-THC was predicted to be a mouse noncarcinogen based on the possession of six fragments (Figure 4.12). Each of the six fragments was found in three other compounds in the mouse learning set, all of which were mouse noncarcinogens (Table 4.23). Δ9-THC is therefore predicted to have a 100% probability of being a mouse noncarcinogen.

Table 4.23 Fragments from the ABC 3/0.90 model leave-one-out validation analysis used to predict the activity of the mouse noncarcinogen 1-trans-delta-9-tetrahydrocannabinol.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag2456	0	3	3	0.000	1.000
Frag2465	0	3	3	0.000	1.000
Frag2474	0	3	3	0.000	1.000
Frag2481	0	3	3	0.000	1.000
Frag2489	0	3	3	0.000	1.000
Frag2539	0	3	3	0.000	1.000
	Probability of activity			0.000	1.000

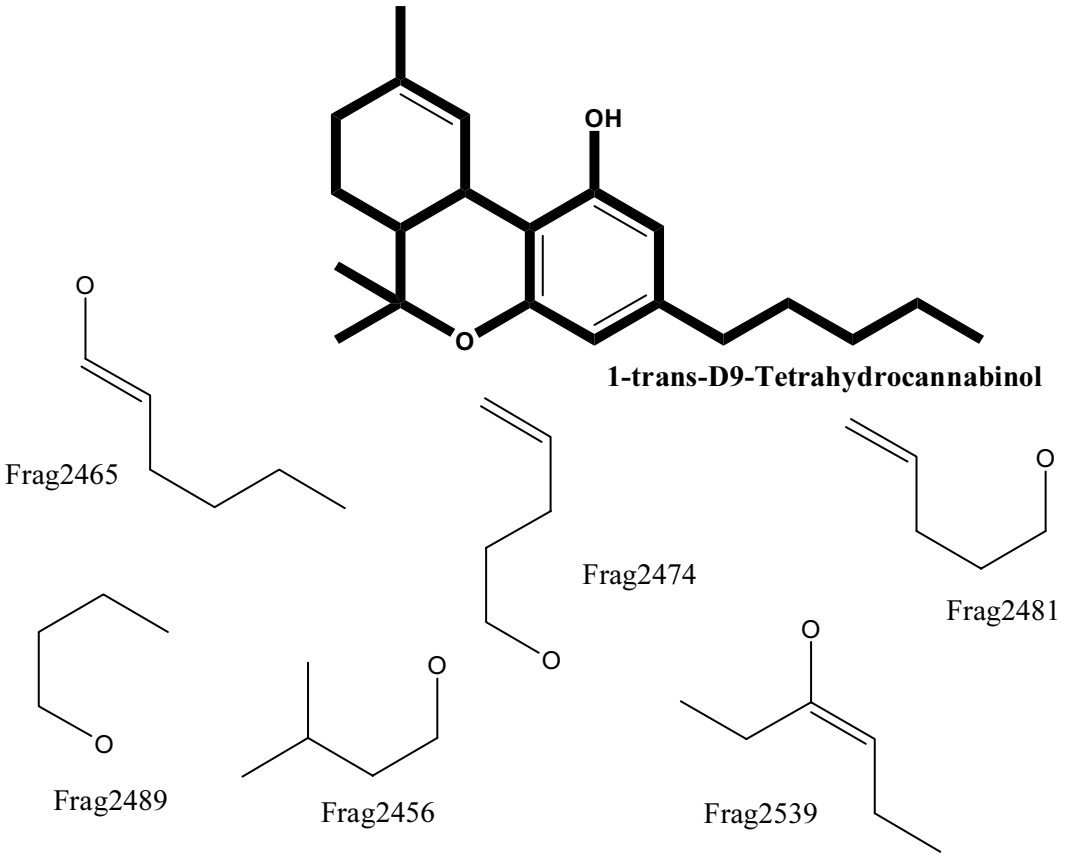

Figure 4.12 Illustration of the 6 significant fragments contributing to the inactive validation prediction of the mouse noncarcinogen 1-trans-delta-9-tetrahydrocannabinol (THC).

4.1.6.11 Acetohexamide

Acetohexamide (Dymelor™) was selected to illustrate cat-SAR prediction of rat carcinogens based on the rat carcinogenesis model. This compound has been tested in male and female rats and mice and has not been observed to induce cancer in any of the four groups. Also, acetohexamide is not classified as a *Salmonella* mutagen. Dymelor™ is a popular sulfonylurea (SUR) drug in commerce used in the treatment of diabetes (an oral antidiabetic drug). There are currently 6 sulfonylureas on the market in the U.S. and almost 1200 SURs in existence. Acetohexamide like other sulfonylurea drugs (i.e., tolbutamide, chlorpropamide, and tolazamide) are first-generation sulfonylureas. In comparison to first-generation sulfonylureas, second-generation sulfonylureas (i.e., glyburide, glipizide and glibornuride) have a more non-polar or lipophilic side chain, which results in a marked increase in their hypoglycemic potency.

The sulfate ($-SO_4$) component of this chemical structure appears to be the major component contributing to the overall inactive behavior of acetohexamide and less to do with the ketone group within it structure. The results also suggest that the amino group attached to the sulfate group may play a critical role as to how the chemical may behave. The substituents that seem to enhance hypoglycemic activity are methyl, amino, acetyl, chloro, bromo, methylthio, and trifluoromethyl groups. The benzene ring should contain one substituent, preferably in the *para*-position. Additionally, the group attached to the terminal nitrogen should be of certain size and should impart lipophilic properties to the molecule. The N-methyl are inactive, N-ethyl have low activity, while N-propyl to N-hexyl are most active. Hypoglycemic activity is lost if N-substituent contains 12 or more carbons. Unique structural features include a p-acetyl moiety and a cyclohexyl group on the terminal urea.

The cat-SAR program predicted acetohexamide to be a rat noncarcinogen based on the possession of 20 structural alerts (Figure 4.13). All 20 fragments except fragment 8568 were derived from other rat noncarcinogens. This fragment was identified as the O-C-N group. Fragment 8568 was not present in any other non-mammary carcinogen in the model. Instead, it was found in a total of three mammary carcinogens in the dataset. This fragment based on the total number of fragments in the model was associated with a 100% chance of being active. Interestingly, when compared to other similar significant fragments, namely fragments 11085, 11053, 11066, 11078, 11073, and 11061, the active structural feature 8568 lacks the sulfate group. The significance of this finding is that of the deactivating effect of the sulfate group. Taken this into consideration, the compound was thus predicted to have a 93% probability of being a rat noncarcinogen.

Table 4.24 Fragments from the ABC 3/0.90 model leave-one-out validation analysis used to predict the activity of the rat noncarcinogen acetohexamide.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag8568	3	0	3	1.000	0.000
Frag11041	0	3	3	0.000	1.000
Frag11045	0	3	3	0.000	1.000
Frag11049	0	3	3	0.000	1.000
Frag11052	0	3	3	0.000	1.000
Frag11053	0	3	3	0.000	1.000
Frag11057	0	3	3	0.000	1.000
Frag11060	0	3	3	0.000	1.000
Frag11061	0	3	3	0.000	1.000
Frag11065	0	3	3	0.000	1.000
Frag11066	0	3	3	0.000	1.000
Frag11072	0	3	3	0.000	1.000
Frag11073	0	3	3	0.000	1.000
Frag11077	0	3	3	0.000	1.000
Frag11078	0	3	3	0.000	1.000
Frag11084	0	3	3	0.000	1.000
Frag11085	0	3	3	0.000	1.000
Frag11098	1	8	9	0.111	0.889
Frag11099	0	4	4	0.000	1.000
Frag11100	1	8	9	0.111	0.889
	Probability of activity			0.068	0.932

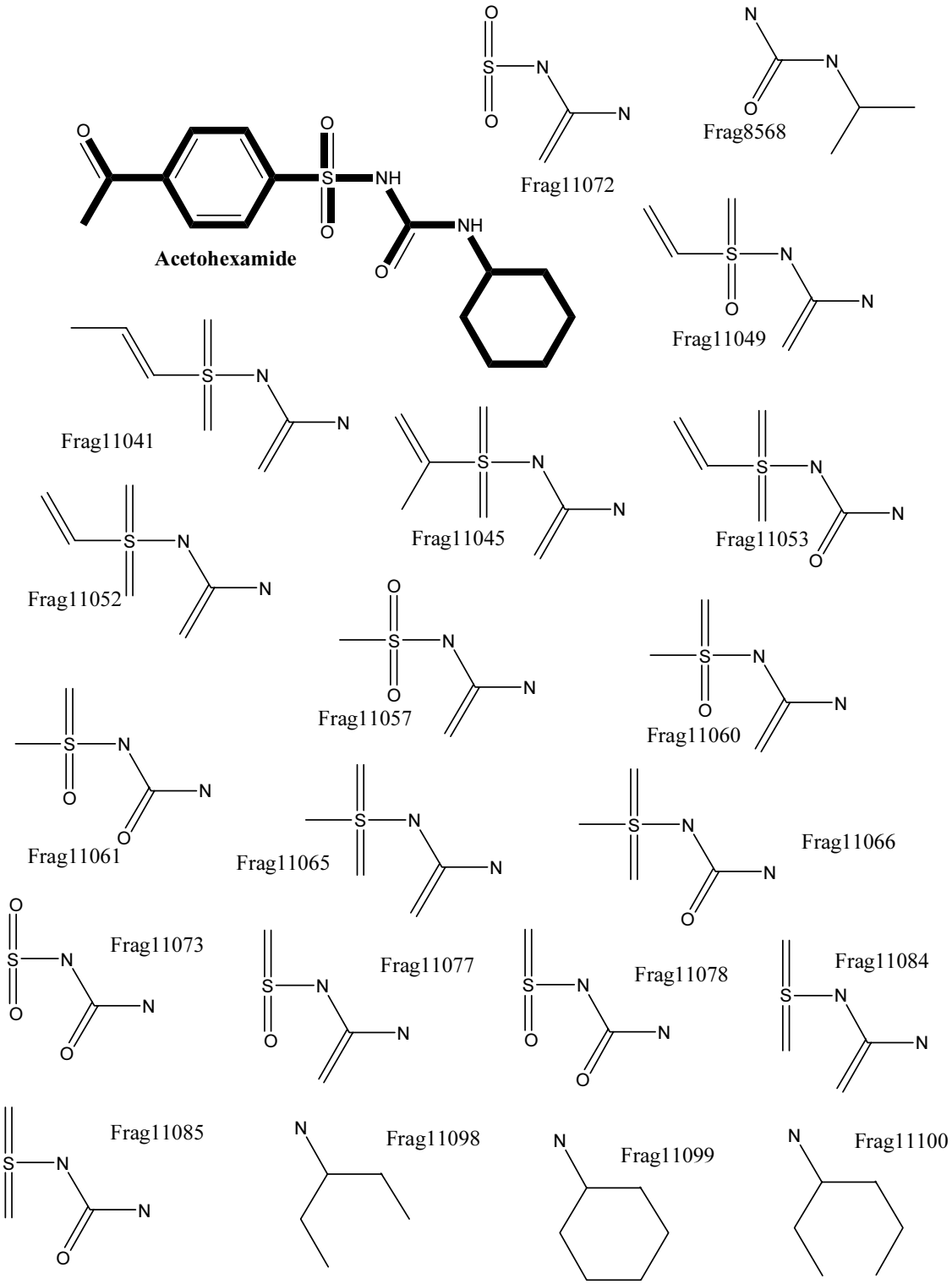

Figure 4.13 Illustration of the 20 significant fragments contributing to the inactive validation prediction of acetohexamide.

4.1.6.12 Bemitradine

Bemitradine was selected to illustrate cat-SAR prediction of rat noncarcinogens based on the rat carcinogenicity model. According to bemitradine's classification in the CPDB, this compound has only been tested in male and female rats. Bemitradine is a diuretic antihypertensive agent that has been shown to cause significant increases in the incidence of liver, thyroid (both sexes), and mammary (females only) neoplasms (Gad et al 1992). The metabolism of bemitradine was studied in both rats and humans. Bemitradine and its primary metabolite (SC-36741; desethylbemitradine) were tested and found to be non-genotoxic in Ames, rat primary hepatocyte UDS, CHO/HGPRT, CHO cytogenetics, in vivo mouse micronucleus and mouse lymphoma TK+/- (bemitradine only) assays (Gad et al 1992). Finally, in an altered hepatic foci (Y-glutamyl transpeptidase positive) promotion assay in female rats, bemitradine was found to be a promoter, though not as potent as phenobarbital. Gad and collaborators concluded that bemitradine (which has been dropped from development) is a non-genotoxic carcinogen that appears to act by a hormonally modulated promotional activity in inducing tumors in the liver and mammary glands. Tumors seen in the thyroid were probably secondary to the effects of bemitradine on metabolism (Gad et al 1992).

The cat-SAR program identified mostly the double presence of the N-C-N atom of the triazine ring to be responsible for the compound's overall carcinogenic effect. Bemitradine was predicted to be a rat carcinogen based on the possession of 10 significant fragments associated with rat carcinogenecity in other compounds in the dataset (Figure 4.14). Each of these fragments was found in three other compounds in the CPDB rat learning

set, all of which were rat carcinogens (Table 4.24). Bemitradine is therefore predicted to have a 100% probability of being a rat carcinogen.

Table 4.25 Fragments from the ABC 3/0.90 model leave-one-out validation analysis used to predict the activity of the rat carcinogen bemitradine.

Fragment	No. Active	No. Inactive	Total	% Active	% Inactive
Frag5998	3	0	3	1.000	0.000
Frag5999	3	0	3	1.000	0.000
Frag6000	3	0	3	1.000	0.000
Frag6009	3	0	3	1.000	0.000
Frag6022	3	0	3	1.000	0.000
Frag6023	3	0	3	1.000	0.000
Frag6024	3	0	3	1.000	0.000
Frag6050	3	0	3	1.000	0.000
Frag6119	3	0	3	1.000	0.000
Frag6145	3	0	3	1.000	0.000
Frag6146	3	0	3	1.000	0.000
Frag6214	3	0	3	1.000	0.000
Frag6215	3	0	3	1.000	0.000
	Probability of activity			1.000	0.000

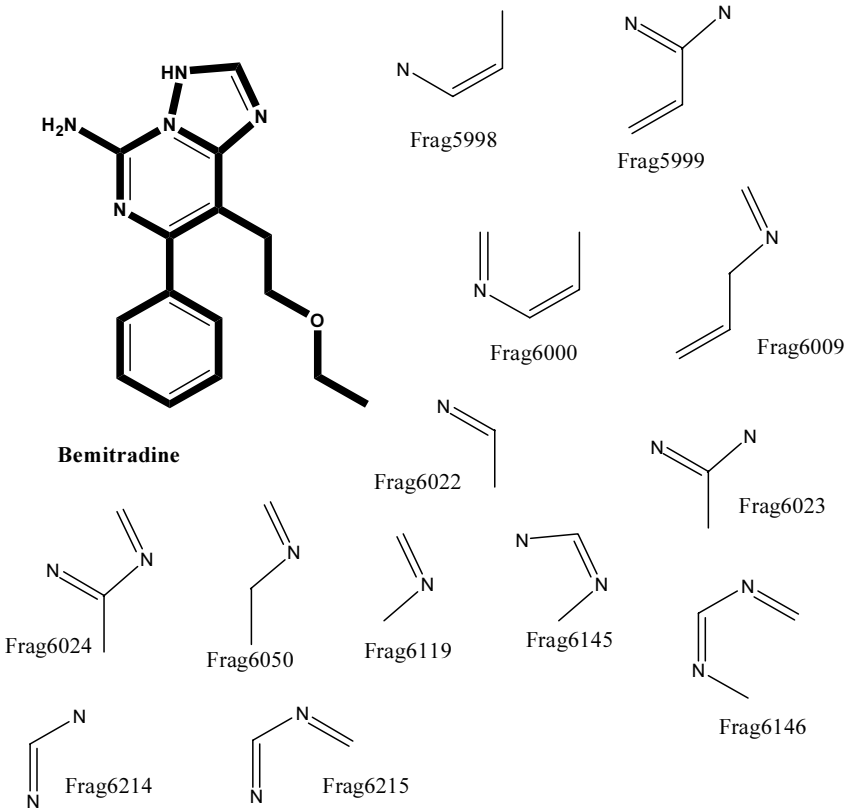

Figure 4.14 Illustration of the 10 significant fragments contributing to the active validation prediction of bemitradine.

4.1.7 Mechanistic Analysis

Analyses of the datasets by the cat-SAR expert system resulted in the derivation of good, explanatory correlations (or the lack thereof) between various toxicological phenomena such as mutagenesis, rodent, female and mammary-specific carcinogenesis and estrogenicity. The CDA was utilized to determine relationships (if any) and common biological effects between the toxicological endpoints studied. The mouse mammary carcinogen models were excluded from the CDA analyses due to it small population of chemicals.

The possible mechanistic overlap between carcinogens, estrogens, and *Salmonella* mutagens was investigated. Based on the CDA analyses, there was a 43.6% correlation between the rat and female rat models (Analysis 1, Table 4.26). This was not unexpected considering the rat female learning set was built from the CPDB rat dataset. It was also reasonable to presume that a positive correlation would be observed between both the rat and female models with the rat mammary carcinogen models for the same reason (i.e., rat MC-NC and MC-NMC models are both subsets of the general rat model). However, the rat mammary models are not exact subsets of the female rat carcinogen model because it included a few male-only breast carcinogens. This, however, should not curtail the model from positively correlating with the female model because only a few chemicals in the mammary carcinogen models were considered male-only breast carcinogens. Hence, the rat MC-NC model had a 72.4% ($p<0.0001$) (Analysis 2, Table 4.26) and 138% (Analysis 3, Table 4.26) correlation with the rat carcinogen and rat female carcinogen models, respectively. Due to this positive correlation observed between these models it was

concluded that there is a relationship between the female rat carcinogen, rat carcinogen, and the rat MC-NC models.

When examining the CDA analyses for similarity of the rat MC-NMC model with other toxicological models the following information was derived. First, the rat MC-NMC model when compared to the rat MC-NC model, correlated to a lesser extent with the rat carcinogen model (19.3%) (Analysis 4, Table 4.26). Secondly, an insignificant ($p=0.032$) correlation was reported when examining the mechanistic overlap between the rat MC-NMC and the female rat carcinogen models (11.4%) (Analysis 5, Table 4.26). This indicates that these two models differ significantly and may be exhibiting their biological effects via different mechanisms. Lastly, the rat MC-NMC model also showed an insignificant ($p=0.010$) relationship to the MC-NC model. There was only a 13.6% similarity between these two models (Analysis 6, Table 4.26) suggesting that these two models are different. This also suggests that there is a difference between the rat MC-NC and MC-NMC models. This was logical considering the rat MC-NMC model is illustrative of the underlying mechanism in which carcinogens are breast carcinogens and the MC-NC model depicts how a substance is carcinogenic.

For the CDA analyses looking at the relationships between *Salmonella* mutagenicity and other carcinogen models, a strong positive and significant correlation was shown. However, the rat MC-NMC and *Salmonella* models were not similar and had a high degree of insignificance associated with it ($p=0.961$). A –0.2% overlap was observed between the two models (i.e., a negative correlation) (Analysis 10, Table 4.26). This non-correlation was explainable. It was speculated that the rat NC-NMC model showed no correlation with the *Salmonella* mutagenicity model due to mutagens being found on both side of the model. In

other words, it was believed that the presence of carcinogens on both sides of the model was responsible for the dissimilarity, and under the model assumption (i.e., dogma) that *Salmonella* mutagens are carcinogens it seems logical that any out-standing electrophiles would be found as they are incorporated into both sides of the model. The *Salmonella* mutagenicity model had a 50.5%, 138.2%, and 128.8% overlap with the rat carcinogen, female rat carcinogen, and rat MC-NC models, respectively (Analysis 7-9, Table 4.26).

Basically, the ESCREEN assay measures estrogen-induced growth of human MCF-7 breast cancer cells. This assay is well characterized and estrogenic responses of chemicals are reported using two unique parameters (i.e., relative proliferative potency (RPP) and relative proliferative effect (RPE)) (Cunningham et al 2004). RPP compares the estrogenic potency of a compound to the potency of the standard estrogen 17-β-estradiol (Cunningham et al 2004). On the other hand, it is realized that many estrogenic compounds, no matter how high the dose, will never produce cell proliferation at the rate of 17-β-estradiol (Cunningham et al 2004). The RPE measures this effect.

When looking at estrogenic relationships to other toxicological phenomena in the CDA analysis, it was shown that estrogens are less likely than random to be *Salmonella* mutagens or carcinogens (Analysis 11-15, Table 4.26). Consistent with the models, *Salmonella* mutagens were not estrogens and estrogens were not carcinogens. It was suggested that since carcinogens overlapped significantly with *Salmonella* mutagens and showed a negative correlation with estrogens that rat carcinogens may be exhibiting their carcinogenic effect mostly through mutagenesis and estrogens might be carcinogenic via some unidentified mechanism. This conclusion was reasonably drawn because the ESCREEN models also failed to denote any correlation with *Salmonella* mutagens.

Interestingly, this outcome was also shown with the MCASE program. It was classically thought that because estradiol and diethylestradiol (DES) carcinogenic mechanisms involved the reduction of electrons, thus producing hydroxyl ions, that estrogens were carcinogens via mutagenesis. This, in fact, might not be the case. The cat-SAR ESCREEN RPE model did not overlap with that of the cat-SAR or MCASE rat and mouse carcinogenicity models. Based on the results, breast carcinogens are not necessarily environmental estrogens and the rat MC-NMC model is finding a unique underlying mechanism of breast carcinogenesis that must be further investigated.

The CDA analysis identifying the common underlying mechanisms (if any) between three toxicological phenomena was also studied. First, the relationship between rat carcinogen, RPE, and *Salmonella* mutagenicity models were examined for mechanistic relatedness. It was found that there were no positive (i.e., -2.3% overlap) or significant ($p=0.743$) relationships between cancer-inducing chemicals in rats, estrogens, and *Salmonella* mutagens (Analysis 16, Table 4.26).

Secondly, the female rat, RPE, and *Salmonella* toxicological endpoints were studied for common mechanistic attributes. Interestingly, although the rat female model was simply a subset of the rat model, a positive correlation was observed when jointly analyzing the rat female carcinogen, RPE, and *Salmonella* mutagenicity models ($p=0.001$). A 31.8% overlap was noted (Analysis 17, Table 4.26). This outcome could not be explained. The rat MC-NC, RPE, and *Salmonella* models had a significant ($p <0.0001$) overlap of 44.1% (Analysis 18, Table 4.26). There was a significant negative correlation (i.e., -75.4%) observed between the MC-NMC, RPE, and *Salmonella* models (Analysis 19, Table 4.26).

Overall, the cat-SAR program is working well. As evidence, the well-tested compounds, caffeine and DES, were reported by cat-SAR as inactive compounds for *Salmonella* mutagenicity and for the ESCREEN RPE model caffeine was inactive, but DES based on its chemical structure was found to be positive for estrogenicity. This suggests that our models are able to distinguish between estrogens and carcinogens. Moreover, an additional CDA analysis of the various toxicological endpoints with *Salmonella* non-mutagens, although not shown in the table, presented just the opposite of the *Salmonella* mutagens. In other words, a negative correlation was noted for each endpoint.

Additionally, an analysis of the *Salmonella* mutagenicity data on active and inactive categories of the rat MC-NC and MC-NMC models was conducted to determine the degree of mutagenicity associated with each. In light of the active categories, 61 of the 73 rat MC-NC and MC-NMC carcinogens evaluated for *Salmonella* mutagenicity were mutagenic (i.e., 83.6%). These results were consistent with the findings by Gold and collaborators who observed 79% of mutagens to be carcinogens and 49% of non-mutagens as carcinogens. In other words, mutagens are more likely to be carcinogenic than nonmutagens.

When analyzing the mutagenicity of the "inactive" category of the rat MC-NMC model (i.e., NMC), it was observed that 57.3% of the model's 75 carcinogens (or nonmammary carcinogens) evaluated in the *Salmonella* assay were mutagenic. However, when evaluating the inactive (i.e., NC) category of the rat MC-NC model, 21% of the model's 66 noncarcinogens were mutagenic. This suggests that when comparing the "inactive" categories of the rat MC-NC and MC-NMC models, a higher degree of mutagenicity is associated with the nonmammary carcinogens (i.e., rat MC-NMC model).

Table 4.26 Mechanistic relationships of the cat-SAR rat mammary carcinogen models and other rodent carcinogenicity, *Salmonella* mutagenicity, and estrogenicity models.

Analysis	Observed	Expected	Δ*	100Δ/Expected	p-value
1. Rat + F-Rat	1166	812	354	43.6	<0.0001
2. Rat + Rat MC-NC	1410	818	592	72.4	<0.0001
3. F-Rat + Rat MC-NC	1202	505	697	138.0	<0.0001
4. Rat + Rat MC-NMC	1335	1119	216	19.3	<0.0001
5. F-Rat + Rat MC-NMC	769	690	79	11.4	0.032
6. Rat MC-NC + Rat MC-NMC	791	696	95	13.6	0.010
7. Rat + Salm	1537	1021	516	50.5	<0.0001
8. F-Rat + Salm	1648	692	956	138.2	<0.0001
9. Rat MC-NC + Salm	1595	697	898	128.8	<0.0001
10. Rat MC-NMC + Salm	933	935	-2	-0.2	0.961
11. Rat + ESCREEN RPE	1165	1284	-119	-9.3	0.010
12. F-Rat + ESCREEN RPE	484	793	-309	-38.9	<0.0001
13. Salm + ESCREEN RPE	684	1094	-410	-37.5	<0.0001
14. Rat MC-NC + ESCREEN RPE	446	799	-353	-44.2	<0.0001
15. Rat MC-NMC + ESCREEN RPE	431	1092	-661	-60.5	<0.0001
16. Rat + RPE + Salm	388	397	-9	-2.3	0.743
17. F-Rat + RPE + Salm	323	245	78	31.8	0.001
18. Rat MC-NC + RPE + Salm	356	247	109	44.1	<0.0001
19. Rat MC-NMC + RPE + Salm	83	337	-254	-75.4	<0.0001

Notes:
Observed: Number of compounds simultaneously identified to be estrogens using the RPE model and the row-listed endpoint.
Expected: The product of the individual prevalences of compounds identified to be estrogens using the RPE model and the row-listed endpoint.
p-value: Difference of two means test.
Δ: Difference of observed from expected.
100Δ/Expected: Percent difference from expected.

4.1.7.1 Similarities in Mechanistic Relatedness: cat-SAR and MCASE

Upon evaluating the CDA analyses of the cat-SAR models, it was noted that the model results for various toxicological phenomena were consistent with that of the widely accepted commercial program MCASE. This served as a great indicator that our models were indeed capable of making valid predictions and model interpretations. The MCASE and cat-SAR *Salmonella* mutagenicity models presented the highest degree of similarity. The results showed a 103.4% overlap occurring between the MCASE and cat-SAR *Salmonella*

models (Analysis 3, Table 4.27). Additionally, significant correlations were observed between MCASE and cat-SAR models for the following endpoints: rat carcinogenesis (28.8%), mouse carcinogenesis (89.1%), ESCREEN RPE (65.7%), and RPP (31.2%) models (Analysis 1,2,4, and 5, Table 4.27).

Table 4.27 Comparison of measures of mechanistic similarities between cat-SAR models of CPDB rat and mouse carcinogenicity, ESCREEN RPE and RPP and *Salmonella* mutagenicity, and MCASE models of the same toxicological phenomena.

Analysis	Observed	Expected	Δ	100Δ/Expected (%)	p-value
1. ESCREEN relative proliferative potency (RPP)	1068	814	254	31.2	<0.0001
2. ESCREEN relative proliferative effect (RPE)	827	499	328	65.7	<0.0001
3. *Salmonella* mutagenicity	2054	1010	1044	103.4	<0.0001
4. CPDB rat	1623	1260	363	28.8	<0.0001
5. CPDB mouse	1199	634	565	89.1	<0.0001

Notes: See Table 4.26

4.2 Discussion

4.2.1 Data Interpretation

It is of great importance to note that the reproducibility of approximately 75%, as attained by Gold and collaborators in the rodent bioassay, may be the upper limit of accuracy reachable by any SAR model. In other words, regardless of how many times the model is evaluated it is never expected to be more than 75% accurate. From this perspective, given the mechanistic complexity of the carcinogenicity phenomena, the cat-SAR CPDB rat and mouse models that were 70% predictive is a respectable performance. The validation results for such models reflected a reasonably high prediction and explanatory power.

It was also observed that some of the test compounds that underwent evaluations for mammary gland carcinogenicity in rats were identified as aromatic amines (e.g., hexamethylmelamine, *o*-toluidine, N-methylaniline), some of which were monocyclic aromatic amines. For example, the chemical *o*-toluidine is a mammary carcinogen in female rats. It may also be fair to say that this particular class of compounds may be exerting its

effect through the same mechanism. However, based on the mammary databases, not all chemicals within a specific chemical class induced mammary gland cancer. The mammary carcinogen database also included some halogenated and epoxide-forming chemicals. This is of particular significance because the pathogenesis of mammary gland neoplasia in the rat is similar to that in humans (Dunnick et al 1995).

Although the described method relied heavily on the fragments, some of which were identified as structural alerts, the overall accuracy of the model was comparable to that of biologically based models. The study results clearly provide a plausible structural explanation as to not only why certain chemicals may be carcinogenic or non-carcinogenic, but also why they may be tissue or organ-specific. In addition, this study did not support the hypothesis that environmental estrogens may play a role in the etiology of mammary gland cancer. This was a significant finding as several studies within recent years have hypothesized that endocrine disruptors and particularly xenoestrogens are etiologic factors in an increased incidence of breast cancer (Lickley et al 2000; Hunter et al 1997; Pieter van't Veer et al 1997; Falck et al 1992; Wolff et al 1995; Wasserman et al 1976). Synthetic chemicals with estrogenic activity (xenoestrogens) and organochlorine environmental contaminants polychlorinated biphenyls (PCBs) and DDE have been the prime etiologic suspects. In other words, estrogen-receptor-mediated mechanisms may be responsible for breast cancer development. However, the results from this study suggests otherwise. It was noted that only a few (i.e., less than 5%) chemicals comprising the mammary carcinogen datasets were estrogenic compounds. This may explain the lack of correlation between the estrogenicity and breast carcinogenicity endpoints. In contrast, strong correlation was observed between *Salmonella* mutagenicity and breast carcinogenicity. In other words, the

chemicals identified as mammary carcinogens in rodents were frequently *Salmonella* mutagens.

In support of this significant finding, Safe *et al.* noted that the results of a large number of studies from several countries on levels of PCBs (total) and DDE in breast cancer patients versus controls clearly demonstrate that levels of DDE and total PCBs are not significantly higher in breast cancer patients versus controls (Safe 2004). Furthermore, many of these studies contained large number of patients, and it was evident that differences in PCB and/or DDE levels in breast cancer patients versus controls initially reported in small cohorts in Connecticut (Falck et al 1992) and New York (Wolff et al 1993) were not subsequently observed in other states/countries (Safe 2004). Gammon and others whose studies focused on the potential linkage of PCBs and DDE and breast cancer in breast cancer patients in Long Island, New York, concluded that "these findings, based on the largest number of samples analyzed to date among primarily white women, do not support the hypothesis that organochlorines increase breast cancer risk among Long Island women" (Gammon et al 2002) (Safe 2004). Hence, the endocrine disruptor hypothesis regarding the induction of mammary gland cancer suggests that exposure to weak environmental estrogens does not play a role in the development of breast cancer.

4.2.2 Strengths and Weaknesses of cat-SAR

The strength of cat-SAR lies in its capacity to find the link between predictions and the reproducibility of rodent cancer bioassays and short-term assays, even in the case of unknown relationship. The data analyses unveil the usefulness of this approach in differentiating between carcinogens and noncarcinogens, and mammary carcinogens and non-mammary carcinogens. This research confirms the feasibility of cat-SAR for predicting

carcinogenicity of chemicals. This expert system's characteristic of being able to identify not just the active fragments of a chemical contributive to its carcinogenic effect but its inactive fragments contributing to a 'noncarcinogenic' effect may not be new to the real world, but it is new the field of computational toxicology and SAR modeling. This makes it possible to study the underlying mechanistic pathways of breast cancer and anti-cancer.

The models developed using our new program's methodology has proven to work on levels equivalent to and in some cases, better than some other existing SAR programs. Our approach can usefully complement the results of the rodent bioassay. A common weakness of the cat-SAR expert system was that it incorrectly identified some very well known inactive compounds as active. Moreover, a few of the most popular powerful carcinogens were predicted to be inactive. At times, it was observed that particular groups or fragments of a test compound shared the same properties of a class of compounds that were not shown by that particular compound in its entirety. Another weakness uncovered in the program was the exclusion of possible structural alerts. These groups may have been associated with aromaticity, electrophilicity and nucleophilicity. It is possible that there may have been a significant group influencing the chemical's overall activity. However, this fragment group may have been unique with respect to the entire group of fragments in the data set and thus, as a rule, excluded from the final fragment set. As a result, the reported list of structural alerts may be incomplete. This may account for some of the wrong active or inactive chemical predictions. Most pitfalls surrounding the cat-SAR algorithm can be linked to the lack of information and structural or model uncertainties (Walker 2003) arising from the fact that every model is a simplification of reality.

Additionally, cat-SAR faces a challenge commonly faced by expert systems. This phenomenon is the effect of neighboring groups. An identified structural alert may be exhibiting carcinogenic effect because of another structural feature on the compound. For any expert system it is a challenge to assess this phenomenon.

CHAPTER 5. SUMMARY AND CONCLUSIONS

5.1 Summary

The cat-SAR expert system was evaluated for its ability to predict and distinguish CPDB mammary-specific carcinogens and rodent carcinogens from non-mammary carcinogens and whole animal noncarcinogens. The validation trials were conducted using the LOO-CV procedure. The mammary carcinogen models developed herein improved SAR models in general, for predicting mammary carcinogenesis by considering chemicals that were carcinogens, but not mammary gland carcinogens, in the dataset. The study results demonstrate the usefulness of cat-SAR in predicting chemical toxicity for datasets of tested and untested compounds, with strength in its ability to predict mutagenicity over a wide range of chemical space, i.e. universal predictivity. The analyses of the predictions indicate that the concordance is equitably weighted in terms of the program's ability to correctly predict mammary carcinogens and non-mammary carcinogens, whereas other SAR approaches tend to produce rodent data of uneven distribution of noncarcinogens favoring a significantly high specificity compared to its sensitivity (i.e., high number of false negatives). Additionally, the performance of the cat-SAR program in predicting CPDB mammary-specific carcinogens appeared to be independent of the chemical utility of the compound, i.e., industrial, agricultural, or pharmaceutical.

The rat and mouse mammary gland carcinogen models were an estimated 80% accurate and the general rodent models were both 70%. Given the fact that predictive models will never achieve 100% overall accuracy because the animal bioassay results from which they are built are 75% accurate, it is reasonable to say that the cat-SAR program is performing at a respectable level. It was observed that the ABCH 3/0.90 models tend to be

more inclusive in terms of the number of test compounds predicted correctly. However, in comparison to the ABC 3/0.90 model for each dataset, the ABCH 3/0.90 models were more accurate overall in terms of its ability to predict the correct activity. Due to the small size of the mouse mammary carcinogen datasets it was presumed that this greatly influenced the favoring of the least restrictive model (i.e., ABC 3/0.75). Thus, it was reasonable to conclude that for large learning sets the following concept applies: the more restricted the model, the more predictive and accurate the model. However, for small datasets such as that of the mouse mammary carcinogen models, a less restrictive (i.e., ABC 3/0.75) model is preferred.

5.2 Conclusions

The application of the cat-SAR program has been demonstrated for outlining the performance characteristics of its predictive SAR models. It provided a rigorous and precise computational basis for making predictions and can be used by governmental agencies as well as industry as a starting point for predicting toxicity and ranking chemicals based on prediction results. The program has proven successful in use of similarity analysis from which an understanding of biological mechanisms that underlie adverse biological effects in humans and ecologically important species can strengthen. The cat-SAR program's capability to identify organ-specific carcinogens makes it a unique and promising SAR approach. Moreover, the methodology presented herein could readily be expanded to other biological endpoints.

The cat-SAR study showed that certain structure-activity relationships exist among mammary and non-mammary gland carcinogens. Several structural alerts believed to activate or deactivate mammary carcinogenesis have been identified. It was found that the amino azo

group was the most prominent structural feature among the mammary carcinogens. Furthermore, the results of this study suggest that there is no linkage between dietary or environmental estrogenic compounds and breast cancer. The underlying relationship between estrogenicity and carcinogenicity is intriguing and merits more detailed investigation, since it may provide new insights into the mechanisms that control mammary tumorigenesis.

Lastly, the structural alerts identified were consistent in randomly selected noncarcinogen models derived from subsets of the rodent data suggesting confidence in an association between these parameters and mammary carcinogenesis. It should be noted that although the results reported herein indicate that structural features of certain mammary gland carcinogens contribute to their ability to induce mammary gland cancer in rodents, the model may not be relevant to human breast carcinogenicity as the data were derived from rodent data, rather than from human studies.

The goal of accurate and reliable toxicity prediction for any chemical, based solely on structural information remains elusive. To predict, we must be able to rely on the accuracy of animal bioassays and generalize based on structural alerts. Evaluation of new SAR methods is needed due to the limitations on the current toxicity screening methods and heightening demands on regulatory processes. The challenge for the future will be to improve technologies for prediction within the constraints of available data, make optimal use of new test data, and better integrate elements of quantitative modeling (QSAR), empirical association, and biological and chemical mechanisms towards the goal of toxicity prediction (Richard 1998). SAR modeling of biological end-points is an empirical study and further research is required to determine the factors associated with the increasing incidence of breast cancer.

There are challenges involved in identifying and comprehending tissue-specific carcinogens. We have tested the rodent and mammary carcinogen models for its predictivity with the LOO-CV procedure and found it to be promising. The cat-SAR method is practical and it is to be used as a tool to assist in medical treatment of breast cancer. It is also believed that cat-SAR can be used in pinpointing those chemicals that really need attention. Therefore, prioritizing them for more costly and time-consuming *in vivo* assays required for the classification of carcinogens.

Moreover, the cat-SAR rodent study is based on a broad spectrum of chemicals of various classes. As such, the models can be generalized based upon cross-validations. Lastly, the structural alerts identified suggest that DNA adduct formations and prohibitions of cell division may be important steps in breast cancer development. Further studies are needed to elucidate the mechanisms underlying estrogen-mediated influences in breast cancer development.

5.3 Suggestions for Future Research Work

Future research work, based on existing work, include, but not limited to the following: (1) Upon extensive review of the CPDB plot used in this study, it was observed that the mammary gland carcinogens for the rat and mouse were established by just a few *in vivo* experiments whose results provided the possibility of it being carcinogenic with respect to the mammary gland. It is likely that more animal cancer tests could in fact improve the quality and consistency of the data produced by enriching the learning set with new mammary carcinogens leading to the identification of new structural features. So theoretically, providing the cat-SAR expert system with the necessary data could improve model predictivity and enhance mechanistic understanding, however, in reality, due to the

high expense and time constraints associated with rodent cancer bioassays, this is very unlikely; (2) further studies into why a test compound was incorrectly predicted by the cat-SAR program as active or inactive should be pursued. Exploring this area could contribute to possible chemical and biological mechanisms; and (3) a distinction should be made between marginally active carcinogens and active carcinogens. The models evaluated in this study require further investigation and critical scrutiny.

REFERENCES

Ames, B.N. and Gold, L.S. (1990). Chemical carcinogenesis: Too many rodent carcinogens. *Medical Sciences*. 87:7772-76.

Ames, B.N., Gold L.S. and Willett, W.C. (1995). The causes and prevention of cancer. *Proc National Academy Science.* 92(12): 5258-65.

Ames B.N., Durston W.E., Yamasaki E., Lee F.D. (1973). Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proc Natl Acad Sci* USA 70:2281-2285.

Andersen, R. Andersson A.M., Arnold S.F., Autrup H., Barfoed M., Beresford N.A., Bjerregaard, P. (1999). Comparison of Short-term Estrogenicity Tests for Identification of Hormone-Disrupting Chemicals. *Environmental Health Perspectives.* 107(1):89-108.

Ankley G.T., Bradbury S.P., Mekenyan O., Ivanov J., Karcher W. and Karabunarliev S. (1997). A Computational-Based Hazard Identification Algorithm That Incorporatess Ligand Flexibility identification of Potential Androgen Receptor Ligands. *Environmental Science and Technology*. 31(12):3702-3711.

Ashby J., Houthoff E., Kennedy S.J., Stevens J., Bars R., Jekat F.W., Campbell P., Van Miller J., Carpanni F.M., Randall G.L. (1997). The challenge posed by endocrine-disrupting chemicals. *Environmental Health Perspectives* 105:164-169.

Ashby J., Elsby R., Sumpter J.P., Brooks A.N., Pennie W.D., Maggs J.L., Lefevre P.A., Odum J., Beresford N., Paton D. and Park K. (2000). Obstacles to the Prediction of Estrogenicity from Chemical Structure: Assay-Mediated metabolic Transformation and the Apparent Promiscuoous Nature of the Estrogen Receptor. *Biochemical Pharmacology*. 60: 1519-1530.

Ashby J. (1994). Two million rodent carcinogens? The role of SAR and QSAR in their detection, *Mutation Research*. 305: 3-12.

Ashby, J. and Paton, D. (1993). The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures. *Mutation Research.* 286: 3-74.

Ashby, J. and Tennant, R.W. (1994). Prediction of rodent carcinogenicity for 44 chemicals: results. *Mutagenesis*. 9(1): 7-15.

Apostolakis E.M., Garai J., Lohmann J.E., Clark J.H., and O'Mally B.W. (2000). Epidermal growth factor activates reproductive behavior independent of ovarian steroids in female rodents. *Molecular Endocrinology.* 14(7):1086-98.

Barratt, M.D. (2000). Prediction of toxicity from chemical structure. *Cell Biology and Toxicology.* 16:1-13.

Barratt, M.D. (1998). Integration of QSAR and *in vitro* Toxicology. *Environmental Health Perspectives*. 2:459-465.

Bartkova J., Lukas J., Muller H., Lutzhoft D., Strauss M., Bartek J. (1994). Cyclin D1 protein expression and function in human breast cancer. *Int J Cancer.* 57(3):353-361.

Benigni, R. and Romano, Z. (2004). The second National Toxicology program comparative exercise on the prediction of rodent carcinogenicity: definitive results. *Mutation Research.* 566:49-63.

Benigni, R. (1991). QSAR prediction of rodent carcinogenity for a set of chemicals currently bioassayed by the US National Toxicology Program. *Mutagenesis.* 6: 423-425.

Bennett M and Davis B.J. (2002). The identification of mammary carcinogens in rodent bioassays. *Environmental and Molecular Mutagenesis*. 39(2-3):150-157.

Birnbaum L.S. and Fenton S.E. (2003). Cancer and developmental exposure to endocrine disruptors. *Environmental Health Perspectives* 111(4):389-94.

Brinton L.A. and Schairer C. (1993). Estrogen replacement therapy and breast cancer risk. *Epidemiol Rev* 15:66-79

Bristol, D.W. (1995). Summary and recommendations: activity classification and structure activity relationship modeling for human health risk assessment of toxic substances. *Toxicology Letters* 79: 265-280.

Brown N.M., and Lamartiniere C.A. (1995). Xenoestrogens alter mammary gland differentiation and cell proliferation in the rat. *Environmental Health Perspectives* 103:708-13.

Brusick D.J. (1994). An assessment of the genetic toxicity of atrazine: relevance to human health and environmental effects. *Mutation Research* 317:133-144.

Butterworth B.E., Popp J.A., Conolly R.B. and Goldsworthy T.L. (1992). Chemically induced cell proliferation in carcinogenesis. Mechanisms of Carcinogenesis in Risk Identification: International Agency for Research on Cancer (IARC) pp. 279-305.

Carlini EA. (1995). The good and the bad effects of (-) trans-delta-9-tetrahydrocannabinol (Delta 9-THC) on humans. *Drug Des Discov.*13(2):155-66.

Chanda S, Erexson G, Riach C, Innes D, Stevenson F, Murli H, Bley K. (1993). Genotoxicity studies with pure trans-capsaicin. *Mol. Pharmacol*. 44, 1008-1015

Colton T. and Greenberg E.R. (1993). Breast cancer in mother prescribed diethylstilbestrol. *Journal of the American Medical Association* 269:2096-2100.

Coyle Y.M. (2004). The effect of environment on breast cancer risk. *Breast Cancer Research and Treatment.* 84:273-288.

Crisp, T.M., Clegg, E.D., Cooper, R.L., Wood, W.P., Anderson, A.G., Baetcke, K.P., Hoffmann, J.L., Morrow, M.S., Rodier, D.J., Schaeffer, J.E., Touart, L.W., Zeeman, M.G., Patel, Y.M. (1998). Environmental endocrine disruption: an effects assessment and analysis. *Environmental Health Perspectives* 106(1):11-56.

Cronin M.T.D., Jaworska J.S., Walker J.D., Comber M.H., Watts C.D., Worth A.P. (2003). Use of QSARs in International Decision-Making Frameworks to Predict Health Effects of Chemical Substances. *Environmental Health Perspectives*. 111:10.

Cunningham A.R., Cunningham S.L., Karol M.H. (accepted 2005). Development of an information-intensive structure-activity relationship model and its application to human respiratory chemical sensitizers. *SAR and QSAR in Environmental Research.* 16(3); revised

Cunningham A.R. and Rosenkranz H.S. (2001). Estimating the extent of the health hazard posed by high-production volume chemicals. *Environmental Health Perspectives*. 109:953-956.

Cunningham A. R.; Cunningham, S. L., and Rosenkranz H. S. (2004). Structure-Activity Approach to the Identification of Environmental Estrogens: The MCASE Approach *SAR and QSAR in Environmental Research* 5(1):55-68.

Cunningham A.R., Klopman G. and Rosenkranz H.S. (1998). Identification of structural features and associated mechanisms of action for carcinogens in rats. *Mutation Research*. 405:9-28.

Cunningham A.R., Rosenkranz H.S., Klopman G., and Zhang Y.P. (1998). Identification of 'genotoxic' and 'non-genotoxic' alerts for cancer in mice: the Carcinogenic Potency Database. *Mutation Research.* 398(1-2):1-17.

Davies T.S. and Monro A. (1995). The case for an upper dose limit of 1000 mg/kg in rodent carcinogenicity tests. *Cancer Letters*. 95: 69-77.

Davis D.L. and Bradlow H.L. (1995). Can Environmental Estrogens Cause Breast Cancer? *Scientific American.*

Davis D.L., Bradlow H.L., Wolff M., Woodruff T., Hoel D.G., and Culver H. (1993). Medical Hypothesis: Xenoestrogens As Preventable Causes of Breast Cancer. *Environmental Health Perspectives* 101:372-377.

Dearden J.C., Barratt M.D., Benigni R., Bristol D.W., Combes R.D., Cronin M.T.D., Judson P.N., Payne M.P., Richard A.M., Tichy M., Worth A.P. and Yourick J.J. (1997). The Development and Validation of Expert Systems for Predicting Toxicity. *ATLA* 25:223-252.

Dearden J.C. (2003). *In silico* prediction of drug toxicity. *Journal of Computer-Aided Molecular Design*. 17:119-127.

De los Santos JF, Buchholz T.A. (2000). Carcinogenesis of the male breast. *Current Treatment Options in Oncology* 1:221-227.

Domchek, S. M.; Eisen, A.; Calzone, K.; Stopfer, J.; Blackwood, A.; Weber, B. L. (2003). Application of Breast Cancer Risk Prediction Models in Clinical Practice. *Journal of Clinical Oncology.* 21(4):593-601.

Dunnick J.K., Elwell M.R., Huff J, and Barrett J.C. (1995). Chemically induced mammary gland cancer in the National Toxicology Program's carcinogenesis bioassay. *Carcinogenesis.* vol.16(2):173-179.

Feuer EJ, Wun L-M, Boring CC, et al. (1993). The lifetime risk of developing breast cancer. *Journal of the National Cancer Institute*. 85:892–897.

Easton D, Narod S, Ford D, Steel M. (1994). The genetic epidemiology of BRCA1. Breast Cancer Linkage Consortium. *Lancet* 344:761.

Eldridge J.C., Tennant M.K., Wetzel L.T., Breckenridge C.B., and Stevens J.T. (1994). Factors Affecting Mammary Tumor Incidence in Chlorotriazine-treated Female Rats. *Environmental Health Perspectives* 102(11):29-36.

El-Tanani M.K. and Green C.D. (1997). Interaction between estradiol and growth factors in the regulation of specific gene expression in MCF-7 human breast cancer cells. *Journal of Steroid Biochemistry and Molecular Biology*. 60(5-6):269-76.

Enslein, K. (1988). An overview of structure activity relationships as an alternative to testing in animals for carcinogenicity, mutagenicity, dermal and eye irritation, and acute oral toxicity. *Toxicology and Industrial Health* 4: 479-498.

Enslein, K. (1993). The future of toxicity prediction with QSAR. *In Vitro Toxicology*. 6:163-169.

Enslein, K., Gombar, V.K. & Blake, B.W. (1994). Use of SAR in computer-assisted prediction of carcinogenicity and mutagenicity of chemicals by the TOPKAT program. *Mutation Research* 305: 47-61.

Eriksson L., Jaworska J., Worth A.P., Cronin M.T., McDowell R.M., Gramatica P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of

classifications- and regression-based QSARs. *Environmental Health Perspectives* 111(10):1361-1375.

Erin N, Boyer P.J, Bonneau R.H, Clawson G.A, Welch D.R. (2004). Capsaicin-mediated denervation of sensory neurons promotes mammary tumor metastasis to lung and heart. *Mutation Research* 557(1):85-97.

EPA (2002) "Priority-Setting in the Endocrine Disruptor Screening Program (EDSP)-Background. Washington, DC: Environmental Protection Agency", [http://www.epa.gov/oscpmont/oscpendo/prioritysetting/background.html]

Ettinger B., Friedman G.D., Bush T., Quesenberry C.P. (1996). Reduced mortality associated with long-term postmenopausal therapy. Obstetrics and Gynecology 87:6-12.

Falck F.J., Ricci A.J., Wolfe M.S. Pesticides and polychlorinated biphenyl residues in human breast lipids and their relation to breast cancer. *Archives of Environmental Health* 47:143-146.

Fendrick J.L., Rafaat A.M., Haslam, S.Z. (1998). Mammary gland growth and development from the postnatal period to menopause: ovarian steroid receptor ontogeny and regulation in the mouse. *J Mammary Gland Biol Neopl* 3:7-22.

Gad S.C., Burton E., Chengelis C.P., Levin S., Piper C.E., Oshiro Y., Semler D.E. (1981). Promotional activities of the non-genotoxic carcinogen bemitradine (SC-33643). *Am J Med.* 70(2):361-72.

Gammon M.D., Wolff M.S., Neugut A.I., Eng S.M. (2002). Environmental toxins and breast cancer on Long Island. II. Organochlorine compound levels in blood. *Cancer Epidemiol. Biomarkers Prevent* 11:686–697.

Ganz, P.A. (2001). The role of hormones in breast carcinogenesis: issues of relevance to female childhood cancer survivors. Division of Cancer Prevention and Control Research, Jonsson Comprehensive Cancer Center at UCLA, Los Angeles, Calfornia,. *Med Pediatr Oncology.* 36(5):514-8.

Gaylor D.W. and Gold L.S. (1995). Quick Estimate of the Regulatory Virtually Safe Dose Based on the Maximum Tolerated Dose for Rodent Bioassays. *Regulatory Toxicology and Pharmacology* 22:57-63.

Glatt H., Bartsch I., Christoph S., Coughtrie, M.W., Falany C.N., Hagen M. et al. (1998). Sulfotransferase-mediated activation of mutagens studied using heterologous expression systems. *Chem Biol Interact* 109(1-3):195-219

Gold L.S., Manley N.B., Slone T.H. and Ward J.M. (2001). Compendium of Chemical Carcinogens by Target Organ: Results of Chronic Bioassays in Rats, Mice, Hamsters, Dogs, and Monkeys. *Toxicol. Pathology* 29:639-652.

Gold L.S., Manley N.B., Slone T.H. and Rohrback L. (1999).Supplement to the Carcinogenic Potency Database (CPDB): Results of Animal Bioassays Published in the General Literature in 1993 to 1994 and by the National Toxicology Program in 1995 to 1996. *Environmental Health Perspectives* 107(4):527-533.

Gold L.S., Ames B.N. and Slone T.H. (1998). What Do Animal Cancer Tests Tell Us About Human Cancer Risk?: Overview of Analyses of the Carcinogenic Potency Database. *Drug Metabolism Reviews* 30(2), 359-404.

Gold L.S., Ames B.N., Slone T.H., and Manley N.B. (2002). Misconceptions About the Causes of Cancer. The Fraser Institute. Vancouver, Canada.

Gold L.S., Slone T.H., Manley N.B., and Bernstein L. (1991). Target Organs in Chronic Bioassays of 533 Chemical Carcinogens. *Environmental Health Perspectives.* 93:233-246.

Gold L.S., Sawyer C.B., Magaw R., Backman G.M., deVeciana M., Levinson R., Hooper N.K., Havender W.R., Bernstein L., Peto R., Pike M.C. and Ames B.N. (1984). A carcinogenic potency database of the standardized results of animal bioassays. *Environmental Health Perspectives* 58:9-319.

Gottmann E., Kramer S., Pfahringer B., and Helma C. (2001). Data Quality in Predictive Toxicology: Reproducibility of Rodent Carcinogenicity Experiments. *Environmental Health Perspectives* 109(5):509-514.

Greenberg E.R., Barnes A.B., Resseguie L. and Barrett J.A. (1984). Breast cancer in mothers given diethylstilbestrol in pregnancy. *New England Journal of Medicine* 311:1393-1398.

Gruber, C.J., Tschugguel W., Schneeberger C., and Huber J.C. (2002). Production and actions of estrogens. *New England Journal of Medicine.* 346:340-52.

Hall, J. M. and Korach, K. S. (2002). Analysis of the Molecular Mechanisms of Human Estrogen Receptors alpha and beta Reveals Differential Specificity in Target Promoter Regulation by Xenoestrogens. *Journal of Biological Chemistry.* 277(46): 44455-44461.

Haslam S.Z. and Woodward T.L. (2003). Host microenvironment in breast cancer development: Epithelial-cell-stromal-cell interactions and steroid hormone action in normal and cancerous mammary gland. *Breast Cancer Research* 5(4):208-215.

Haworth S., Lawlor T., Mortelmans K., Speck W. and Zeiger E. (1983). Salmonella Mutagenicity Test Results for 250 Chemicals. *Environmental Mutagenesis.* 1:3-42.

Heneweer M., Muusse M., Dingemans M., de Jong P.C., van den Berg M., and Sanderson J.T. (2004). Co-Culture of Primary Human Mammary Fibroblasts and MCF-7 Cells as an in Vitro Breast Cancer Model. *Toxicol Science*

Hileman, B. (1994). Environmental estrogens linked to reproductive abnormalities, cancer. *Chem Eng News.* 72:19-23.

Hulzebos, E.M. and Posthumus, R. (2003). (Q)SARs: Gatekeepers Against Risk On Chemicals? *SAR and QSAR in Environmental Research.* 14(4):285-316.

Hunter D.J. (1997). Plasma Organochlorine Levels and the Risk of Breast Cancer. *New England Journal of Medicine* 337(18):1253-1258.

Jadeski L.C., Chakraborty C., and Lala P.K. (2003). Nitric oxide-mediated promotion of mammary tumour cell migration requires sequential activation of nitric oxide synthase, guanylate cyclase and mitogen-activated protein kinase. *Int J Cancer* 106(4):496-504.

Jaworska J.S., Comber M., Auer C., and Van Leeuwen C.J. (2003). Summary of a Workshop on Regulatory Acceptance of (Q)SARs for Human Health and Environmental Endpoints. *Environmental Health Perspectives.* 111(10): 1358-1360.

Jobling S., Reynolds T., White R., Parker M.G., Sumpter J.P. (1995). A variety of environmentally persistent chemicals, including some phthalate plasticizers, are weakly estrogenic. *Environmental Health Perspectives* 103:582-587.

Johnson, F.M. (2003). How many high production chemicals are rodent carcinogens? Why should we care? What do we need to do about it? *Mutation Research.* 543:201-215

Karmali RA, Volkman A, Muse P, Louis TM. (1979). The influence of diazepam administration in rats bearing the R3230AC mammary carcinoma. *Prostaglandins Med* 3:193-198.

Klaassen, C.D. (2001). Casarett and Doull's Toxicology: The Basic Science of Poisons. 6th Edition. McGraw Hill 4:86-87.

Kleinerman RA, Brinton LA, Hoover R, Fraumeni JF Jr. (1981). Diazepam and breast cancer. *Lancet* 1:1153.

Kleinerman RA, Brinton LA, Hoover R, Fraumeni JF Jr. (1984). Diazepam use and progression of breast cancer. *Cancer Res* 44:1223-1225.

Klopman, G. (1984). Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. *Journal of the American Chemical Society.* 106: 7315-7320.

Klopman, G. (1992). Multi-CASE: a hierarchical computer automated structure evaluation program. *Quantitative Structure-Activity Relationships* 11: 176-184.

Klopman, G. and Rosenkranz, H.S. (1994). Approaches to SAR in carcinogenesis and mutagenesis. Prediction of carcinogenicity/mutagenicity using Multi-CASE. *Mutation Research* 305: 33-46.

Kuroki Y., Fukushima K., Kanda Y., Mizuno K., Watanabe Y. (2000). Putative membrane-bound estrogen receptors possibly stimulate mitogen-activated protein kinase in the rat hippocampus. *European Journal of Pharmacology.* 400(2-3):205-209.

Lee SM, O'Connor PJ, Thatcher N, Crowther D, Margison GP, Cooper DP. (1994). Relationships between the formation of O6-methyldeoxyguanosine by 1-p-carboxyl-3,3-dimethylphenyltriazene in DNA and O6-alkylguanine-DNA alkyltransferase in human peripheral leukocytes. *Cancer Research* 54(15):4072-6.

Le Marchand L., Haiman C.A., Wilkens L.R., Kolonel L.N., and Henderson B.E. (2004). MTHFR Polymorphisms, Diet, HRT, and Breast Cancer Risk: The Multiethnic Cohort Study. *Cancer Epidemiol Biomarkers Prev*. 13(12):2071-7.
Lickley D.R., Fish L.A., Hiraki E.B., Holloway G.Y., Ross C., Hanna T., SenGupta W.M., Weber S.K. (2000). Breast adipose tissue concentrations of polychlorinated biphenyls and other organochlorines and breast cancer risk. *Cancer Epidemiol. Biomarkers Prevent* 9:55–63.

Linusson A., Gottfries, J., Lindgren F., and Wold, S. (2000). Statistical Molecular Design of Building Blocks for Combinatorial Chemistry. *J. Med. Chem.* 43:1320-1328.

Lundkvist, U. (1990). Clinical and reproductive effects of Clophen A50 (PCB) administered during gestation on pregnant guinea pigs and their offspring. *Toxicology.* 61:249-257.

Lowis, D.R. A New, Highly Predictive QSAR Technique. (1997). *Tripos Technical Notes* 1(5):1-7.

Malacarne, D. Pesenti, R., Paolucci, M. & Parodi, S. (1993). Relationship between molecular connectivity and carcinogenic activity: a confirmation with a new software program based on graph theory. *Environmental Health Perspectives* 101: 332-342.

Markey C.M., Luque E.H., Munoz de Toro M.M., Sonnenschein C., Soto A.M. (2001). *In utero* exposure to bisphenol A alters the development and tissue organization of the mouse mammary gland. *Biology of Reproduction*. 65:1215-1223.

Jaworska J.S. (2002). Bayesian Analyses and Inference from QSAR Predictive Model Results. *SAR and QSAR in Environmental Research.* 13(1): 111-125.

Martin A.M. and Weber B.L. (2000). Genetic and hormonal risk factors in breast cancer. *Journal of National Cancer Institute* 92:1126-1135.

Mekenyan O., Kamenska V., Serafimova R., Poellinger L., Brouwer A., and Walker J. (2002). Development and Validation of an Average Mammalian Estrogen Receptor-Based

QSAR Model. *SAR and QSAR in Environmental Research.* 13(6): 579-595.

Melvin L.S., Milne G.M., Johnson M.R., Wilken G.H., Howlett A.C. (1996). Structure-activity relationships defining the ACD-tricyclic cannabinoids: cannabinoid receptor binding and analgesic activity. *J Immunol.*157(10):4412-20.

Miller E.C., Swanson A.B., Phillips D.H., Fletcher T.L., Liem A, Miller J.A. (1983). Structure-activity studies of the carcinogenicities in the mouse and rat of some naturally occurring and synthetic alkenylbenzene derivatives related to safrole and estragole. *Cancer Research* 43(3):1124-34.

Miller BA, Kolonel LN, Bernstein L, et al. (eds). (1996). *Racial/Ethnic Patterns of Cancer in the United States 1988–1992*, National Cancer Institute. NIH Pub. No. 96–4104. Bethesda, MD

Moysich K.B., Menezes R.J., Baker J.A. and Faulkner K.L. (2002). Environmental Exposure to Polychlorinated Biphenyls and Breast Cancer Risk. *Reviews on Environmental Health.* 17(4):263-277.

Mueller, S.O. (2004). Xenoestrogens: mechanisms of action and detection methods *Analytical and Bioanalytical Chemistry* 378(3):582-587.

Mueller, S. O.; Kling, M.; Arifin Firzani, P.; Mecky, A.; Duranti, E.; Shields-Botella, J.; Delansorne, R.; Broschard, T.; Kramer, P. J. (2003). Activation of estrogen receptor a and ERb by 4-methylbenzylidene-camphor in human and rat cells: comparison with phyto- and xenoestrogens. *Toxicology Letters.* 142(1-2):89-101.

Nelson, H.D. (2002). Assessing Benefits and Harms of Hormone Replacement Therapy. *Scientific Review and Clinical Applications* 288(7):882-884.

National Toxicology Program (NTP). (1978). Bioassay of formulated fenaminosulf for possible carcinogenicity (CAS No. 140-56-7). Natl Cancer Inst Carcinog Tech Rep Ser.; 101:1-105.

National Toxicology Program (NTP). (1996). NTP Toxicology and Carcinogenesis Studies of 1-Trans-Delta(9)-Tetrahydrocannabinol (CAS No. 1972-08-3) in F344 Rats and B6C3F1 Mice (Gavage Studies). Natl Toxicol Program Tech Rep Ser.; 446:1-317.

Pai S.B. (1983). Evaluation of mutagenic effect of the fungicide femaminosulf in Drosophila melanogaster. *Mutation Research* 119(3):289-91.

Palmer JR, Hatch EE, Rosenberg CL, Hartge, Kaufman RH, Titus-Ernsto L, Noller K.L., Herbst A.L., Rao R.S., Troisi R., Colton T., Hoover R.N. (2002). Risk of breast cancer in women exposed to diethylstilbestrol in utero: Preliminary results (United States). *Cancer Causes and Control.* 13: 753-758.

Peto R, M.C. Pike, L. Bernstein, L.S. Gold and Ames B.N. (1984). *Environmental Health Perspect*ives. 58: 1-8.

Pike A, Brzozowski A.M., Hubbard R.E. and Bonn T. (1999). Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. *The EMBO Journal.* 18:4608-4618.

Pilot, H.C., III, and Dragan Y.P. Chemical Carcinogenesis. 1996. Chapter 8 in Casarett and Doull's Toxicology, 5th Edition, Curtis D. Klassen, Mary O. Amdur, and John Doull, Eds., McGraw-Hill, New York, pp. 201-268.

Pollack N., Cunningham A. R., and Rosenkranz H. S. (2003). Environmental persistence of chemicals and their carcinogenic risks to humans. *Mutation Research.* 528(1-2):81-91.

Recchia, A. G.; Vivacqua, A.; Gabriele, S.; Carpino, A.; Fasanella, G.; Rago, V.; Bonofiglio, D.; Maggiolini, M. (2004). Xenoestrogens and the induction of proliferative effects in breast cancer cells via direct activation of oestrogen receptor alpha. *Food Additives and Contaminants* 21(2):134-144.

Richard, A.M. (1999). Application of artificial intelligence and computational methods to predicting toxicity. *Knowledge Eng. Rev.* 14:307-317.

Richard A.M. (1998). Commercial toxicology prediction systems: a regulatory perspective. *Toxicol Letter* 102-103:611-6.

Richard A.M. (1998). Structure-based methods for predicting mutagenicity and carcinogenicity: are we there yet? *Mutation Research.* 25; 400(1-2):493-507.

Richard, A.M. (1994). Application of SAR methods to non-congeneric databases associated with carcinogenicity and mutagenicity: issues and approaches. *Mutation Research* 305: 73-97.

Ries LAG, Eisner MP, Kosary CL, et al. (eds). *SEER Cancer Statistics Review, 1975–2001*, National Cancer Institute. Bethesda, MD, 2004 (http://seer.cancer.gov/csr/1975_2001).

Rosenkranz, H.S. Cunningham A.R. and Pollack N. (2001). Exploring the relationship between the inhibition of gap junctional intercellular communication and other biological phenomena. *Carcinogenesis* 21(5):1007-1011.

Rosenkranz, H.S. and Cunningham, A.R. (2001). SAR Modeling of Unbalanced Data Sets. *SAR and QSAR in Environmental Research* 12(3):267-274.

Rosenkranz, H.S. and Rosenkranz, M. (1997). Intercellular communication, tumor promtion and non-genotoxic carcinogenesis: relationships based upon structural considerations. *Mutation Research.* 381:171-188.

Roy D., Floyd R.A., Liehr J.G. (1991). Elevated 8-hydroxyguanine levels in DNA of diethylstilbestrol-treated Syrian hamsters: Covalent DNA damage by free radicals generated by redox cycling of diethylstilbestrol. *Cancer Research* 51

Rudel RA, Brody JG, Spengler JD, Vallarino J, Geno PW, Sun G, Yau A (2001). Methods to detect selected potential mammary carcinogens and endocrine disruptors in commercial and residential air and dust samples. *Journal of the Air and Waste Management Association.* 51:499-513.

Russo, I.H. and Russo, J. (1996). Mammary Gland Neoplasia in Long-Term Rodent Studies. *Environmental Health Perspectives* 104:938-967.

Safe S.H. (2004). Endocrine disruptors and human health: is there a problem. *Toxicology* 205:3-10.

Safe S.H. (2002). Environmental estrogens: roles in male reproductive tract problems and in breast cancer. *Rev Environ Health.* 17(4): 249-52.

Safe S.H. (1995). Environmental and Dietary Estrogens and Human Health: Is there a Problem? *Environmental Health Perspectives* 103(4):346-351.

Setchell, K.D.R. Naturally Occurring Non-Steroidal Estrogens of Dietary Origin. In Estrogens in the Environment. J.A. McLachlan, ed. New York: Elsevier Science Publishing Co., Inc., 1985, pp. 69-85.

Shi L. M.; Fang, H.; Tong, W.; Wu J.; Perkins R.; Blair R. M.; Branham W. S.; Dial S. L.; Moland C. L.; Sheehan D. M. (2001). QSAR Models Using a Large Diverse Set of Estrogens. *Journal of Chemical Information and Computer Sciences* 41(1):186-195.

Simoncini T. and GenazzaniA. (2003). Non-genomic actions of sex steroids hormones. *European Journal of Endocrinology* 148:281-292

Singh, A. K. (2001). Development of QSAR Models to Predict Estrogenic, Carcinogenic, and Cancer Protective Effects of Phytoestrogens *Cancer Investigation* 19(2):201-216.

Singh S, Natarajan K, Aggarwal B.B. (2004). Capsaicin (8-methyl-N-vanillyl-6-nonenamide) is a potent inhibitor of nuclear transcription factor-kappa B activation by diverse agents. *Anticancer Research* 24(2B):1003-9.

Skillman T.G. and Feldman J.M. (2004).The pharmacology of sulfonylureas. *Toxicon.* 44(4):461-7.

Snyder R.D., Pearl G.S., Mandakas G., Choy W.N., Goodsaid F., and Rosenblum I.Y. (2004) Assessment of the Sensitivity of the Computational Programs DEREK, TOPKAT, and MCASE in the Prediction of the Genotoxicity of Pharmaceutical Molecules. *Environmental and Molecular Mutagenesis*. 43:143-158.

Soto A.M., Chung K.L. and Sonnenschein C. (1994). The Pesticides Endosulfan, Toxaphene, and Dieldrin Have Estrogenic Effects on Human Estrogen-Sensitive Cells. *Environmental Health Perpectives*. 102(4):380-383.

Stadel B.V. (2002). Hormone replacement therapy and risk of breast cancer. *JAMA* 287:2360-2361; author reply 2361.

Starek, A. (2003). Estrogens and organochlorine xenoestrogens and breast cancer risk. *International Journal of Occupational Medicine and Environmental Health* 16(2):113-124.

Steinmetz R., Young P.C.M., Caperell-Grant A., Gize E.A., Madhukar B.V., Ben-Jonathan N., and Bigsby R.M. (1996). Novel Estrogenic Action of the Pesticide Residue β–Hexachlorocyclohexane in Human Breast Cancer Cells. *Cancer Research*. 56:5403-5409

Stoica G.E., Franke T.F., Wellstein A., Czubayko F., List H.J., Reiter R., Morgan E., Martin M.B., Stoica A. (2003). Estradiol rapidly activates Akt via the Erbβ2 signaling pathway. *Molecular Endocrinology*. 17(5):818-30.

Stoll BA. (1976). Psychosomatic factors and tumour growth. In: Risk Factors in Breast Cancer, (Stoll BA, ed) London:William Heinemann Medical Books; 193-203.

MacDonald D.J. and Luker M.A. (1980). Fluoride: interaction with chemical mutagens in Drosophila. *Mutation Research* 71(2):211-218

Tennant R.W. (1993). *Mutation Research* 286:111-118

Tomatis L. and Huff J. (2001) Evolution of cancer etiology and primary prevention. *Environmental Health Perspectives* 109:5-7

Tong W, Lowis D.R., Perkins R, Chen Y, Welsh W.J., Goddette D.W., Heritage T.W. and Sheehan D.M. (1998). Evaluation of quantitative structure-activity relationship methods for large-scale prediction of chemicals binding to the estrogen receptor. *J Chem Inf Computer Science* 38(4):669-77.

Pieter van't Veer. (1997). DDT (dicophane) and postmenopausal breast cancer in Europe: case-control study. *British Journal of Medicine* 315:81-85.

Vogel E. (1973). Strong antimutagenic effects of fluoride on mutation induction by trenimon and I-phenyl-3,3-dimethyltriazene in *Drosophila melanogaster*.
Mutation Research Vol. 3:339-52.

Wagner, P.M., Nabholz, J.V., Kent, R.J. (1995). The new chemicals process at the Environmental Protection Agency (EPA): Structure-activity relationships for hazard identification and risk assessment. *Toxicol. Let.* 79:67-73.

Wakeling A.E., Valcaccia B., Newboult E. and Green L.R. (1984). Non-steroidal Antioestrogens-Receptor Binding and Biological response in Rat uterus, Rat Mammary Carcinoma and Human Breast Cancer Cells. *Journal of Steroid Biochemistry* 20(1):111-120.

Walker J.D., Carlson L., and Jaworska J. (2003). Improving Opportunities for Regulatory Acceptance of QSARs: the Importance of Model Domain, Uncertainty, Validity and Predictability. *Quant. Struct.-Act. Relat*. 22:1-5.

Walker J.D., Dimitrov S, Mekenyan O. (2003). Using HPV chemical data to develop QSARs for non-HPV chemicals: opportunities to promote more efficient use of chemical testing resources. *QSAR Combo Sci.* 22:386-395.

Walker J.D., Carlsen L., Hulzebos E. and Simon-Hettich B. (2002). Global Government Applications of Analogues, SARs and QSARs to Predict Aquatic Toxicity, Chemical or Physical Properties, Environmental Fate Parameters and Health Effects of Organic Chemicals. *SAR and QSAR in Environmental Research.* 13(6): 607-616.

Waller C.L., Minor D.L. and McKinney J.D. (1995). Using three-dimensional quantitative structure-activity relationships to examine estrogen receptor binding affinities of polychlorinated hydroxybiphenyls. *Environmental Health Perspectives.* 103:702-207.

White A.C., Mueller R.A., Gallavan R.H., Aaron S., and Wilson A.G.E. (2003). A multiple *in silico* program approach for the prediction of mutagenicity from chemical structure. *Mutation Research.* 539: 77-89.

Wassermann M. (1976). Organochlorine Compounds in Neoplastic and Adjacent Apparently Normal Breast Tissue. *Bulletin of Environmental Contamination and Toxicology* 15(4):478-484.

Wolff M.S. and Toniolo P.G. (1995). Environmental Organochlorine Exposure as a Potential Etiologic Factor in Breast Cancer. *Environmental Health Perspectives* 103(7):141-145.
Wolff M.S. (1993). Blood Levels of Organochlorine Residues and the Risk of Breast Cancer. Journal of the National Cancer Institute 85(8):648:652.

Woo Y.T., Lai D.I., Argus M.F. and Arcos J.C. (1995). Development of Structure Activity Relationship Rules for Predicting Carcinogenic Potential of Chemicals *Toxico Lett.* 79:219-228.

Yu S.J.; Keenan, S. M.; Tong, W.; Welsh, W. J. (2002). Influence of the Structural Diversity of Data Sets on the Statistical Quality of Three-Dimensional Quantitative Structure-Activity Relationship (3D-QSAR) Models: Predicting the Estrogenic Activity of Xenoestrogens. *Chemical Research in Toxicology.* 15(10):1229-1234.

Zacharewski T.R., Fieldin M.R., Matthews J.B., Fertuck K.C. and Halgren R.G. (2002). *In Silico* Approaches to Mechanistic and Predictive Toxicology: An Introduction to Bioinformatics for Toxicologists. *Critical Reviews in Toxicology*. 32(2):67-112.

Zeeman M., Auer C.M., Clements R.G., Nabholz J.V., and Boethlings R.S. (1995). U.S. EPA regulatory perspectives on the use of QSAR for new and existing chemical evaluations. *SAR and QSAR in Environmental Research.* 3(3): 179-201.

Zeiger E., Ashby J., Bakale G., Enslein K., Klopman G., Rosenkranz H.S. (1996). Prediciton of *Salmonella* mutagenicity. *Mutagenesis.* 11:471-84.

Zeiger E. and Piegorsch W.W. (1991). Measuring Intra-assay Agreement for the Ames *Salmonella* Assay. *Statistical Methods in Toxicology*. 43: 35-41.

HPV Challenge Program. http://www.epa.gov/chemrtk/volchall.htm, revision dated 9/18/2001, downloaded 10/18/2001.

NCI Online (Cancer Statistics) available from http://cancer.gov; accessed July 8, 2003.

Gold L.S. Carcinogenic Potency Database (CPDB Online) available from http://potency.berkeley.edu; accessed August 29, 2003.

Congressionally Directed Medical Research Program (CDMRP) data available from http://cdmrp.army.mil/bcrp, accessed May 19, 2003.

VITA

The author was born in beautiful Nassau, New Providence Island, Bahamas. Her family later relocated to Freeport, Grand Bahama Island where she was graduated Valedictorian at a local high school in 1998. Granted a full academic scholarship to study abroad, she further went on to successfully obtain her Bachelor of Science degree in biology from the University of Nebraska in 2002. During her years as an undergraduate, she took up a job at AIM Environmental Consultant Agency as a soil analyst for the Parkinson's disease Etiology Project (PDEP) where she worked for two years under the direction of her major professor. She then decided to continue her studies at the graduate level and inspired by the impressive works of Dr. Albert R. Cunningham in structure-activity relationship modeling, she enrolled in the Environmental Toxicology program in the Department of Environmental Studies where he served as her mentor and research advisor. Ms. Moss is a candidate for a Master of Science degree in the field of environmental toxicology to be awarded in May 2005.